Inhoudsopgave

Voorwoord

Voorwoord bij de vijfde editie

Gedurende vele jaren heb ik het vak statistiek gegeven aan Odisee University College Brussel.
Bij het gebruik van statistische technieken hoort veel rekenwerk. Ervaring leert mij dat de studenten dit, zelfs met een rekenmachine, als vervelend beschouwen. Bovendien is het tijdrovend. Dat laatste is niet onbelangrijk. De voorbije jaren is het aantal college-uren voor het vak statistiek verminderd. Voor dit probleem wil ik een oplossing bieden met dit boek over statistische berekeningen met behulp van het softwarepakket MS Excel. Dit pakket beschikt over een uitgebreid arsenaal aan statistische functies. Die voeren het rekenwerk snel uit. Bovendien zijn de studenten vandaag ook vertrouwd met Excel.

Studenten leren dus statistiek beoefenen met behulp van de pc. Dit houdt in dat we de behandelde statistische onderwerpen eerst uitleggen en dan op de pc toepassen aan de hand van voorbeelden met behulp van de overeenkomstige Excelfuncties. De Excelfuncties zelf worden daarbij stapsgewijs uitgelegd en hun functienamen worden zowel in het Engels als in het Nederlands opgegeven.

De behandelde onderwerpen zijn fundamenteel, zodat het geheel kan worden opgevat als basisleerstof voor het vak statistiek.

In een aparte bijlage achteraan in het boek hebben we oefeningen opgenomen om het inzicht in de materie en in het gebruik van de statistische Excelfuncties te vergroten. Voor de docenten is een handleiding met de uitgewerkte oefeningen beschikbaar.

In deze vijfde editie zijn de procedures voor het uitvoeren van de statistische functies aangepast naar MS Excel 2016.

Ik hoop dat ook deze nieuwe uitgave met succes haar weg zal vinden.

Tot slot dank ik Line De Caluwé en Inge De Winter van Uitgeverij De Boeck voor de aangename en constructieve samenwerking.

dr. Jacques Van Der Elst

hoofdstuk 1

Statistische basisbegrippen

1 Steekproef en populatie

Een steekproef wordt getrokken uit een populatie, om op basis van de resultaten van die steekproef een conclusie te trekken over de populatie.

Voorbeeld
Je wordt gevraagd het stemgedrag te voorspellen bij een referendum over euthanasie in Vlaanderen.
Door een representatieve steekproef van personen die al gestemd hebben te ondervragen, kun je op basis van de verzamelde steekproefresultaten met een bepaalde zekerheid een besluit trekken over het stemgedrag van de populatie inzake euthanasie.

2 Steekproefgrootheden en parameters

Kenmerkende eigenschappen van een steekproef noemen we steekproefgrootheden. Dat zijn bijvoorbeeld het steekproefgemiddelde, de modus, de steekproefstandaardafwijking ...

Karakteristieke eigenschappen van een populatie worden aangeduid met de term parameter. Voorbeelden zijn: het populatiegemiddelde, de populatieproportie ...

3 Discrete en continue variabele

Een discrete variabele is een kenmerk van een element uit een populatie dat slechts bepaalde waarden kan aannemen. Zo'n variabele is bijvoorbeeld het aantal bestuurders in een naamloze vennootschap.

Een continue variabele kan alle reële waarden aannemen. De lengte of het gewicht van een persoon zijn voorbeelden van een continue variabele.

4 Afhankelijke en onafhankelijke variabele

In tegenstelling tot een constante is een variabele een veranderlijke grootheid. Voorbeelden zijn: de koers van een aandeel, de wisselkoers van de dollar, je vermogen enz.

In de statistiek is men zeer geïnteresseerd in het opsporen van verbanden tussen variabelen. Meer bepaald wil men uitmaken of veranderingen in een onafhankelijke variabele ook veranderingen in de afhankelijke variabele teweegbrengen. Bijvoorbeeld: verandert de productiehoeveelheid als men de productiemethode wijzigt?

5 Frequentie

Nemen we het voorbeeld van een steekproef waarbij van 40 studenten van de hogeschool XYZ de uitslag voor het vak economie (op 20 punten) wordt geregistreerd:

14	9	15	17	11
14	10	16	13	18
12	14	13	12	14
18	13	19	12	12
16	8	12	15	16
12	13	11	12	17
20	13	16	10	13
11	12	18	19	12

We kunnen nu een tabel opmaken met twee kolommen. De eerste kolom geeft een overzicht van de mogelijke uitkomsten van de variabele. In de tweede kolom staat voor iedere waarneming het aantal keren dat die waarneming is voorgekomen, of dus de frequentie van iedere observatie.

Uitslag	Frequenties
8	1
9	1
10	2
11	3
12	9
13	6
14	4
15	2
16	4
17	2
18	3
19	2
20	1
	n = 40

Deze tabel toont ons onder meer dat de uitslag 13 een frequentie 6 heeft.

6 Probabiliteit

Als we een muntstuk opwerpen, dan kunnen we ofwel kruis ofwel munt als uitkomst hebben. In de onderstaande tabel geven we de resultaten weer van drie steekproeven met een muntstuk.

	Steekproeven		
Uitkomsten	(1) n = 10	(2) n = 50	(3) n = ∞
Kruis	0,30	0,40	0,50
Munt	0,70	0,60	0,50
	1,00	1,00	1,00

We stellen vast dat bij een groot aantal worpen iedere mogelijke uitkomst van het opwerpen van een muntstuk eenzelfde percentage behaalt. De percentages die we verkrijgen bij een onbeperkt aantal worpen, noemen we de kansen van de mogelijke uitkomsten met een muntstuk.

Mathematisch uitgedrukt betekent dit dat de percentages verkregen bij een groot aantal worpen n, naderen tot limietwaarden die de kansen aangeven van iedere uitkomst:

$$P = \lim_{n \to \infty} \frac{f}{n}$$

waarbij:

f = absoluut aantal keren van een uitkomst
n = aantal worpen

Zo'n limietwaarde noemt men de ***statistische probabiliteit*** van een uitkomst van een toevallig verschijnsel.

hoofdstuk 2

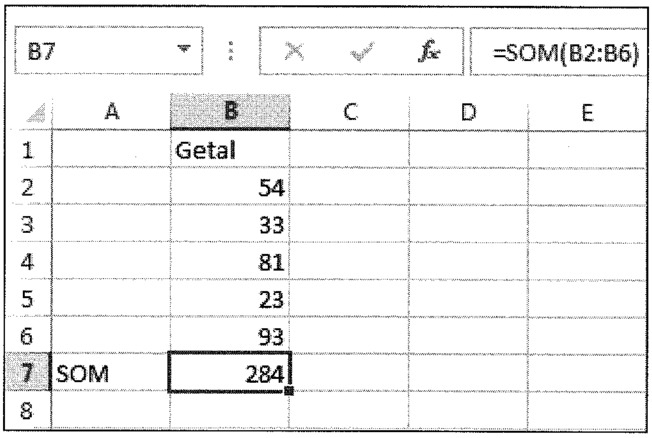
Statistische functies en gegevensanalysefuncties van Excel

1 Werkbladfuncties

Een typische werkbladfunctie is SOM (SUM). Deze functie telt de getallen op in een aantal aangeduide cellen en maakt de som in een andere opgegeven cel.

Voorbeeld

Maak de som van 54, 33, 81, 23 en 93.



B7 =SOM(B2:B6)

	A	B	C	D	E
1		Getal			
2		54			
3		33			
4		81			
5		23			
6		93			
7	SOM	284			
8					

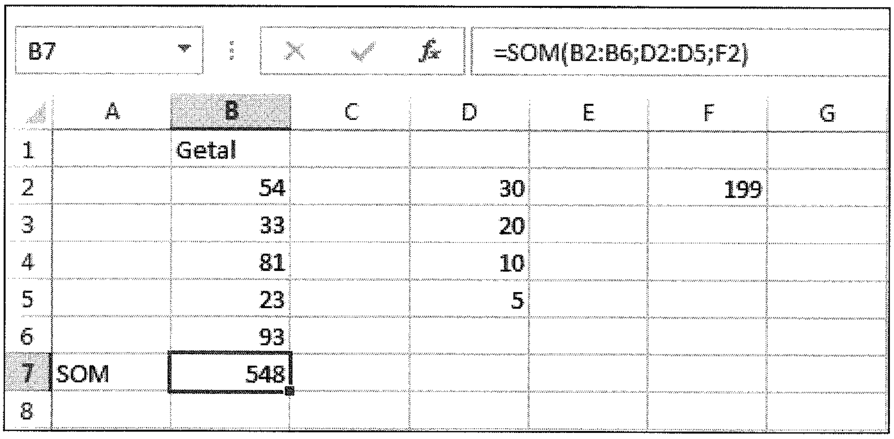
Figuur 1 SOM-functie

De formule in de formulebalk geeft aan dat de waarde in B7 gelijk is aan de som van de getallen opgegeven in de cellen B2 tot en met B6.

SOM is verder een multigegevensfunctie. Dat betekent dat we met SOM gegevens kunnen verwerken die zich eender waar in het werkblad bevinden.

Voorbeeld

Maak de som van het bereik B2:B6, D2:D5 en van F2 in cel B7.



B7 =SOM(B2:B6;D2:D5;F2)

	A	B	C	D	E	F	G
1		Getal					
2		54		30		199	
3		33		20			
4		81		10			
5		23		5			
6		93					
7	SOM	548					
8							

Figuur 2 SOM als multigegevensfunctie

Algemeen gebruiken we een werkbladfunctie als volgt:

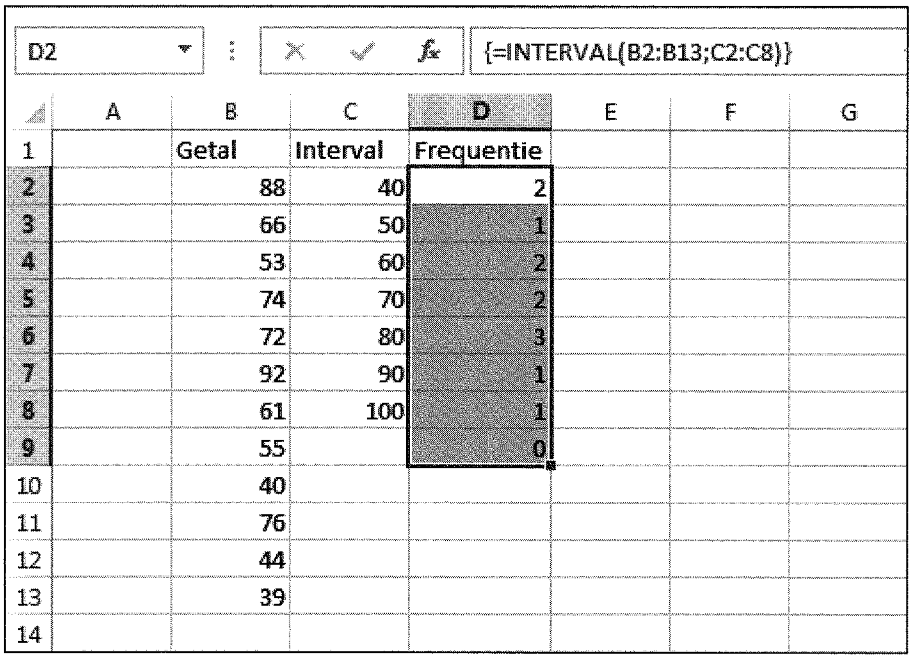
> tik de gegevens in een gegevensbereik en selecteer de cel waarin het resultaat moet komen;
> klik op de f_x-knop om een functie in te voegen;
> selecteer de te gebruiken functie, klik op 'OK' en het dialoogvenster Functieargumenten verschijnt;
> in het dialoogvenster tikken we de vereiste gegevens in;
> klik op 'OK' en het resultaat verschijnt in de geselecteerde cel.

2 Matrixfuncties

Een matrixfunctie berekent meerdere waarden en plaatst deze waarden in een cellenmatrix. De INTERVAL-functie (FREQUENCY) is een voorbeeld van een matrixfunctie. Deze functie rangschikt een verzameling waarden in een reeks vooraf bepaalde intervallen om de frequentie van ieder interval te kennen. Een tabel met intervallen en frequenties levert dan een frequentieverdeling op.

Voorbeeld
In figuur 3 tikken we in het werkblad in het bereik B2:B13 de volgende getallen in: 88, 66, 53, 74, 72, 92, 61, 55, 40, 76, 44, 39.

> Tik de volgende intervallen in de matrix C2:C8: 40, 50, 60, 70, 80, 90 en 100.
> Selecteer het bereik of de matrix voor de frequenties per interval. Omdat INTERVAL een verticale matrix oplevert met een cel meer dan de intervalmatrix, selecteren we D2:D9.
> Klik op de f_x-knop.
> Kies de categorie 'Statistisch' en selecteer dan INTERVAL.
> In het gegevensmatrixvak van het venster Functieargumenten vullen we het gegevensbereik B2:B13 in.
> In het intervalmatrixvak vullen we de matrix C2:C8 in.
> Druk op CTRL + SHIFT + OK om het dialoogvenster te sluiten. De frequentiematrix is nu ingevuld.



D2 {=INTERVAL(B2:B13;C2:C8)}

	A	B	C	D	E	F	G
1		Getal	Interval	Frequentie			
2		88	40	2			
3		66	50	1			
4		53	60	2			
5		74	70	2			
6		72	80	3			
7		92	90	1			
8		61	100	1			
9		55		0			
10		40					
11		76					
12		44					
13		39					
14							

Figuur 3 INTERVAL-functie

3 Gegevensanalysefuncties

Excel beschikt over veel interessante functies voor gegevensanalyse. We vermelden hier enkele technieken die we zullen behandelen: variantieanalyse, correlatie, de F-toets, regressie enz.

Om deze technieken te gebruiken, selecteren we in het lint het tabblad 'Gegevens' en klikken we vervolgens op de knop 'Gegevensanalyse' waarna het dialoogvenster met de functies verschijnt.
Ter kennismaking gaan we even naar de functie 'Beschrijvende statistiek' (Descriptive statistics). Met deze functie kunnen we een aantal steekproefgrootheden berekenen.

Voorbeeld

> We vullen het bereik B2:B11 in met getallen.
> We selecteren 'Gegevensanalyse' via het tabblad 'Gegevens'. Het gegevensanalysedialoogvenster wordt geopend.
> We klikken op de functie 'Beschrijvende statistiek', klikken op 'OK' en het dialoogvenster van 'Beschrijvende statistiek' verschijnt.
> In het Invoerbereikvenster duiden we de gegevensmatrix B1:B11 aan. Merk op dat de coördinaten absoluut zijn.
> Dan drukken we op de Kolom-knop om aan te geven hoe de gegevens gestructureerd zijn.
> We duiden aan dat het Invoerbereik een titel bevat door 'Labels' te selecteren in de eerste rij.

> Klik op de Nieuw tabblad-werkbladknop. Excel maakt dan een nieuw tabblad binnen een werkblad en toont de resultaten in dat nieuwe blad.
> Ten slotte selecteren we 'Samenvattingsinfo' en klikken op 'OK'.

Het uiteindelijke resultaat is een nieuw tabblad met statistische grootheden i.v.m. onze gegevensmatrix.
De onderstaande figuren geven respectievelijk de invoer en de output van de functie 'Beschrijvende statistiek' weer.

	A	B	C
1		Getal	
2		35	
3		53	
4		28	
5		82	
6		33	
7		67	
8		76	
9		44	
10		92	
11		19	
12			

Figuur 4 Invoer voor de Beschrijvende statistiek-functie

	A	B	C
1	*Getal*		
2			
3	Gemiddelde	52,9	
4	Standaardfout	7,939283483	
5	Mediaan	48,5	
6	Modus	#N/B	
7	Standaarddeviatie	25,1062188	
8	Steekproefvariantie	630,3222222	
9	Kurtosis	-1,400637601	
10	Scheefheid	0,270588601	
11	Bereik	73	
12	Minimum	19	
13	Maximum	92	
14	Som	529	
15	Aantal	10	
16			

Figuur 5 Output van de Beschrijvende statistiek-functie

Op de betekenis van de diverse grootheden gaan we later in.

hoofdstuk 3

Grafieken

1 De kolomgrafiek

We nemen aan dat we een commerciële presentatie moeten geven over de omzetevolutie van de diverse productgroepen in een onderneming voor de periode 20X1-20X5. Daartoe stellen we in Excel de onderstaande tabel op.

Tabel 1 Omzetevolutie productgroepen tijdens de periode 20X1-20X5 (mio euro's)

	A	B	C	D	E	F	G	H
1	Tabel 1 Omzetevolutie productgroepen tijdens de periode 20X1-20X5 (mio euro's)							
2								
3		20X1	20X2	20X3	20X4	20X5		
4	Brood	1	1,4	1,4	1,2	1,6		
5	Gebak	0,8	1,4	1,6	1,7	1,9		
6	Pralines	0,1	0,15	0,3	0,35	0,4		
7	Koekjes	0,5	0,3	0,4	0,47	0,6		
8	Totaal	2,4	3,25	3,7	3,72	4,5		

Met een grafiek, bijvoorbeeld een kolommendiagram, kunnen we de trends in de omzetevoluties van de diverse productgroepen gemakkelijk en snel vaststellen.

We tonen hoe je een kolommendiagram kunt maken:

> selecteer de gegevens in A4:F7 en klik op het tabblad 'Invoegen'. Kies vervolgens in de groep 'Grafieken', het grafiektype. Selecteer de keuzelijst 'Kolomdiagram invoegen' waarin je het subtype 'Gegroepeerde kolom' kiest;

> je krijgt onmiddellijk een grafiek te zien die je uiteraard nog moet aanpassen. Selecteer het tabblad 'Indeling', kies in de keuzelijst 'Grafiekgebied' het grafiekelement 'Grafiektitel', typ in de formulebalk 'Omzet productgroepen 20X1-20X5' en klik op de entertoets;

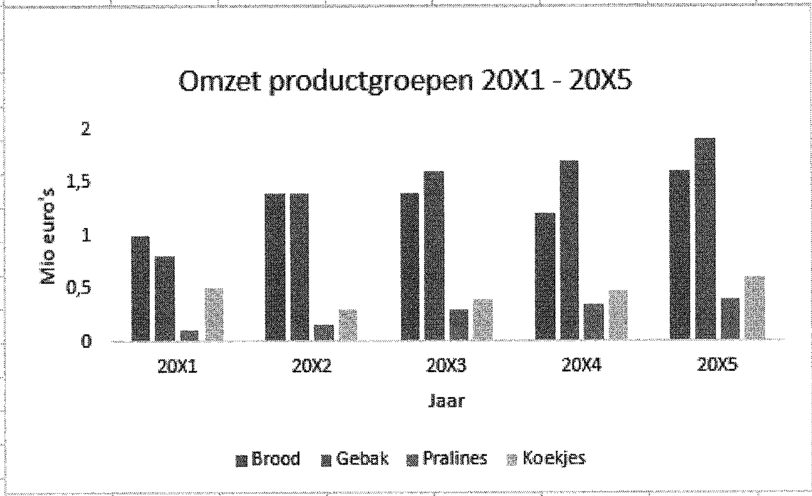
> voeg vervolgens ook de astitels toe. Klik op het tabblad 'Ontwerpen' en kies in de groep 'Grafiekindelingen' de keuzelijst 'Grafiekonderdeel toevoegen'. Kies 'Astitels', selecteer 'Primair horizontaal', typ 'Jaar' in de formulebalk en klik op de entertoets. Voeg ook nog een titel bij de verticale of de waarde-as toe. Selecteer in de keuzelijst 'Grafiekonderdelen toevoegen' opnieuw het grafiekonderdeel 'Astitels'. Klik op 'Primair verticaal', typ 'Mio euro's' in de formulebalk en klik op de entertoets;

> een voorlaatste stap is het vervangen van de horizontale aslabels door de reeksnamen '20X1 tot en met 20X5'. Selecteer het tabblad 'Ontwerpen' en kies in de groep 'Gegevens' de knop 'Gegevens selecteren'. In het venster 'Gegevensbron selecteren' klik je in 'Horizontale aslabels (categorieën)' op 'Bewerken'. Het dialoogvenster 'Aslabels' verschijnt. Voer in het vak 'Aslabelbereik' het gegevensbereik B3:F3 in om het gewenste resultaat te bekomen. Klik ten slotte op 'OK' in de vensters 'Aslabels' en 'Gegevens selecteren';

> tot slot kun je de rasterlijnen nog verwijderen. Selecteer in de keuzelijst 'Grafiekonderdeel toevoegen', 'Rasterlijnen' en kies 'Primair groot horizontaal'. De rasterlijnen zijn nu verdwenen en de grafiek is voltooid. Het resultaat zie je in Figuur 6.

Tabel 1 Omzetevolutie productgroepen tijdens de periode 20X1 - 20X5 (mio euro's)

	20X1	20X2	20X3	20X4	20X5
Brood	1	1,4	1,4	1,2	1,6
Gebak	0,8	1,4	1,6	1,7	1,9
Pralines	0,1	0,15	0,3	0,35	0,4
Koekjes	0,5	0,3	0,4	0,47	0,6
Totaal	2,4	3,25	3,7	3,72	4,5



Figuur 6 Kolomgrafiek en gegevens

2 De cirkelgrafiek

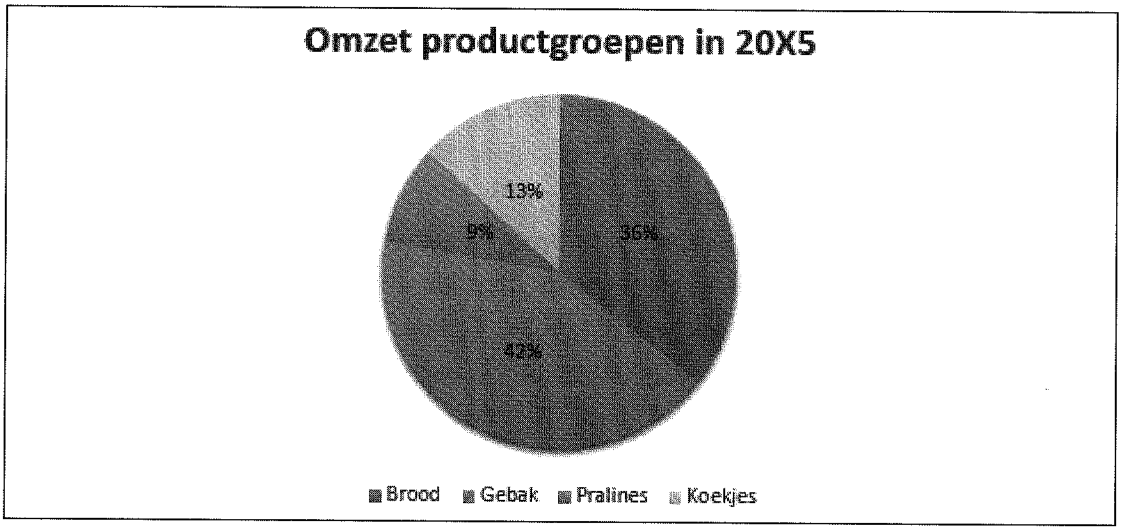
Deze grafiek wordt gebruikt om de procentuele aandelen van de verschillende elementen in het totaal visueel te tonen.

We hernemen de tabel met de omzetevoluties van de verschillende productgroepen en willen het aandeel van elke productgroep in de totaalomzet van 20X5 kennen.
De betreffende cirkelgrafiek maken we als volgt:

> selecteer het gegevensbereik F4:F7;
> klik op het tabblad 'Invoegen' en kies het grafiektype in de groep 'Grafieken';
> selecteer de keuzelijst 'Cirkel- of ringdiagram invoegen' en kies het subtype '2D-cirkel';
> selecteer in de groep 'Grafiekstijlen', 'Stijl 3' zodat het omzetaandeel van de verschillende productgroepen in de totaalomzet van 20X5 verschijnt';
> klik op het tabblad 'Ontwerpen' en kies in de keuzelijst 'Grafiekenonderdeel toevoegen' het grafiekelement 'Grafiektitel' en selecteer 'Boven grafiek'. Typ in de formulebalk 'Omzet productgroepen in 20X5' en klik op de entertoets;

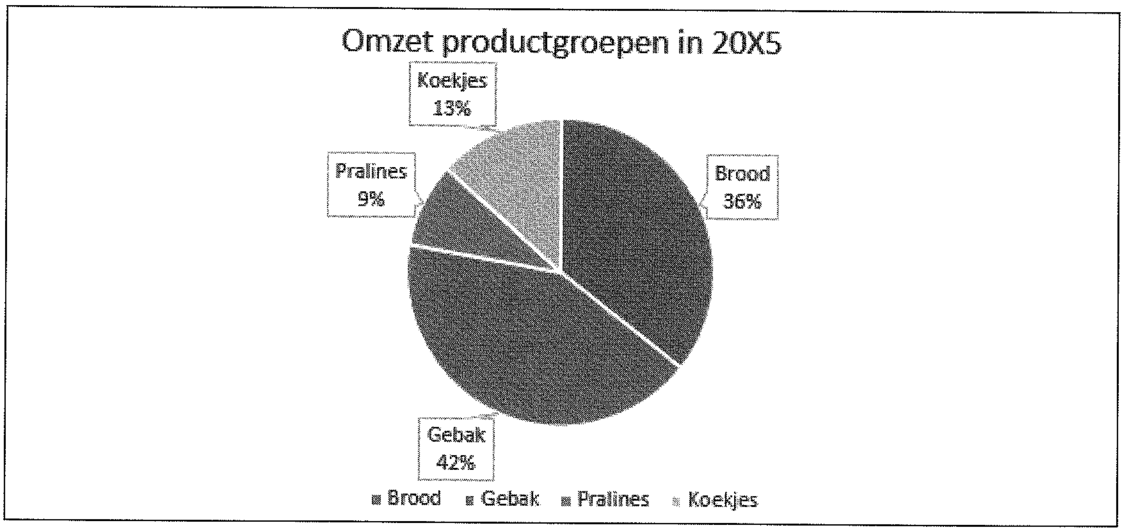
> geef ten slotte nog de werkelijke categorienamen op via de groep 'Gegevens' met de keuzelijst 'Gegevens selecteren'. Het dialoogvenster 'Gegevensbron selecteren' verschijnt. Klik op 'Bewerken' van 'Horizontale labels (categorieën)'. Voer in het vak 'Aslabelbereik' van het venster 'Aslabels' het bereik A4:A7 in en klik op OK in de vensters 'Aslabels' en 'Gegevensbron selecteren'.

Het resultaat is de grafiek in Figuur 7.



Figuur 7 Cirkelgrafiek

Je kunt de grafiek ook laten uiteenvallen in de verschillende segmenten. Klik daarvoor met de linkermuisknop in het grafiekgebied, selecteer het tabblad 'Ontwerpen' en duid in de groep 'Grafiekstijlen' het subtype 'Stijl 1' aan. Met de knop 'Grafiekonderdeel toevoegen' kun je gegevenslabels toevoegen. Selecteer 'Bijschrift bij gegevens' en van elke productgroep verschijnt in de figuur naast het aandeel in de totaalomzet ook de productbenaming. Het resultaat zie je in Figuur 8.

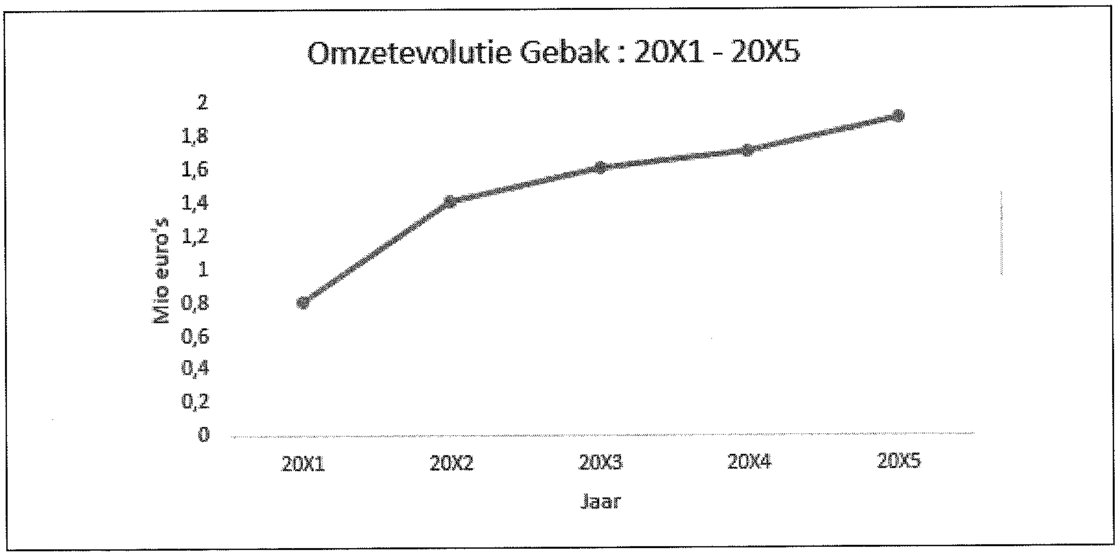


Figuur 8 Uitgelichte cirkelsegmenten

3 De lijngrafiek

De lijngrafiek wordt dikwijls gebruikt om de evolutie van een variabele in de tijd weer te geven. We nemen bijvoorbeeld uit Tabel 1 (p. 25) de productgroep Gebak, waarvan we de omzetevolutie voor de periode 20X1-20X5 lineair willen voorstellen. Om de lijngrafiek te maken, voer je de volgende procedure uit:

> selecteer het gegevensbereik B5:F5, klik op het tabblad 'Invoegen', kies de keuzelijst 'Lijn- of vlakdiagram invoegen' en selecteer 'Lijn met markeringen';
> selecteer het tabblad 'Indeling' en open in de keuzelijst 'Grafiekgebied', 'Grafiektitel', vul in de formulebalk 'Omzetevolutie gebak: 20X1-20X5' in en klik op de entertoets;
> voeg vervolgens ook de astitels toe. Voorzie eerst een titel voor de horizontale of categorie-as met 'Primair horizontaal'. Als titel gebruik je 'Jaar'. Om de verticale as van een titel te voorzien, selecteer je 'Primair verticaal' en vul je 'Mio euro's in';
> om de horizontale aslabels te vervangen door de reeksnamen '20X1 tot en met 20X5' (kolomkoppen), selecteer je het tabblad 'Ontwerpen' en in de groep 'Gegevens' de knop 'Gegevens selecteren'. In het venster 'Gegevensbron selecteren' klik je in 'Horizontale aslabels (categorieën)' op 'Bewerken' en het dialoogvenster 'Aslabels' verschijnt. In het vak 'Aslabelbereik' voer je ten slotte het gegevensbereik B3:F3 in om het gewenste resultaat te bekomen;
> om de rasterlijnen te verwijderen, kies je in de keuzelijst 'Grafiekonderdeel toevoegen' het onderdeel 'Rasterlijnen' en selecteer je de optie 'Primair horizontaal'. Het resultaat zie je in Figuur 9.



Figuur 9 Lijngrafiek

4 De staafgrafiek

Dit grafiektype wordt gebruikt wanneer de categorienamen voor de onafhankelijke veranderlijke x, zijnde de kolom- of rijkoppen van de gegevens op het werkblad, te lang zijn. In dat geval wordt de horizontale as de y-as, en de verticale as de x-as.

We veronderstellen dat de onderstaande tabel een overzicht geeft van het percentage kinderen dat uit elke studierichting van het middelbaar onderwijs in een school in het jaar 20X1 aan clubsport deed.

Tabel 2 Clubsportbeoefening

	A	B	C
1	Tabel 2 Clubsportbeoefening door leerlingen uit een middelbare school in 20X1		
2	Studierichting	Percentages	
3	Kunst	5	
4	Beroeps	12	
5	Algemeen	25	
6	Technisch	30	
7	Sport	70	
8			

De voorstelling van deze gegevens in een staafdiagram verloopt als volgt:

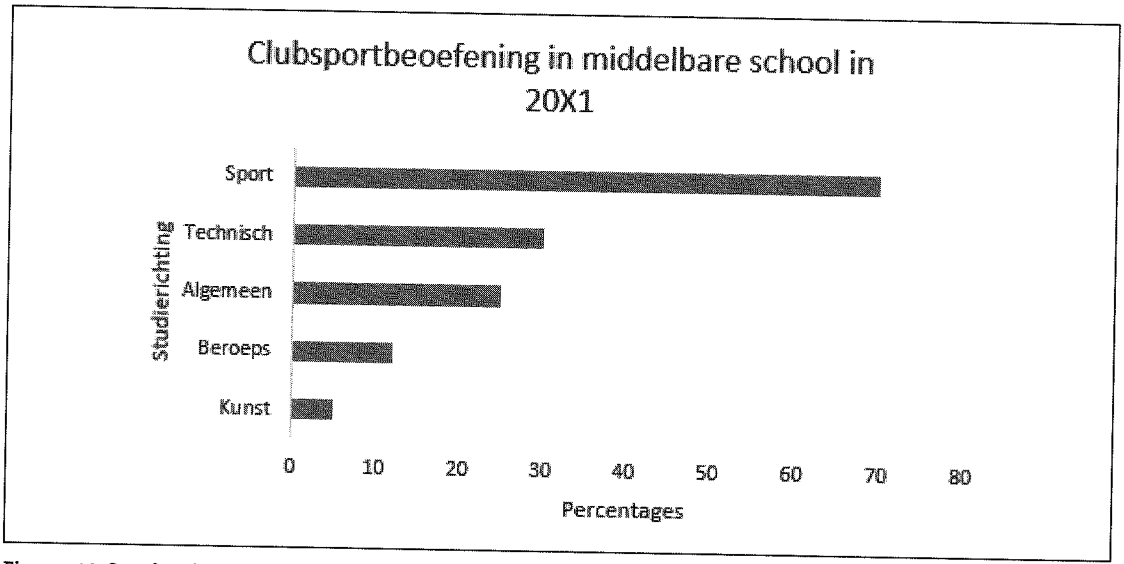
> selecteer het gegevensbereik B3:B7, klik op het tabblad 'Invoegen', kies in de groep 'Grafieken' de keuzelijst 'Kolom- of staafdiagram invoegen' en selecteer 'Gegroepeerde staaf';

> klik op het tabblad 'Indeling' en op de keuzelijst 'Grafiekgebied'. In het vak 'Grafiektitel' vul je 'Clubsportbeoefening in middelbare school in 20X1' in;

> voeg ook de juiste categorienamen toe. Selecteer het tabblad 'Ontwerpen', klik dan in de groep 'Gegevens' op 'Gegevens selecteren'. Het dialoogvenster 'Gegevensbron selecteren' verschijnt. Klik op 'Bewerken' van 'Horizontale labels (categorieën)' en in het vak 'Aslabelbereik' van 'Aslabels' voer je het bereik A3:A7 in dat de juiste categorienamen bevat. Klik op 'OK' in beide vensters om deze stap te beëindigen;

> voeg nu de astitels toe. Klik op het tabblad 'Ontwerpen', ga naar de groep 'Grafiekindelingen' en klik daarna in de keuzelijst 'Grafiekonderdeel' op 'Astitels'. Selecteer 'Primair horizontaal' en vul in het vak 'Astitel', 'Percentages' in. Voeg daarna ook nog een titel toe bij de verticale of waarde-as. Selecteer de keuzelijst 'Grafiekonderdeel toevoegen' opnieuw en klik op 'Astitels'. Kies vervolgens de optie 'Primair verticaal' en typ 'Studierichting';

> verwijder ten slotte nog de verticale rasterlijnen. Selecteer daarvoor in de keuzelijst 'Grafiekonderdeel toevoegen' het onderdeel 'Rasterlijnen' en klik op 'Primair groot verticaal'. Het resultaat zie je in Figuur 10.



Figuur 10 Staafgrafiek

5 Lineaire regressie

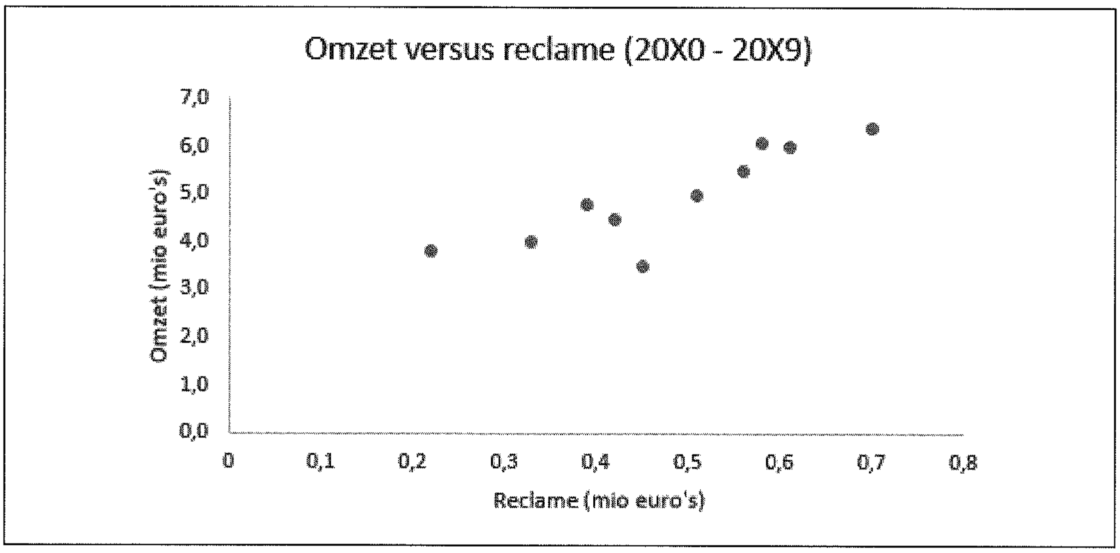
Deze statistische techniek bestudeert het verband tussen twee variabelen. Stel dat we willen nagaan hoe goed de reclame-uitgaven van een onderneming de omzet kunnen voorspellen. Daartoe bekijken we de volgende tabel.

Tabel 3 Reclame-uitgaven en omzetresultaten (20X0-20X9)

	A	B	C
1	Tabel 3 Reclame-uitgaven en omzetresultaten (20X0 - 20X9)		
2			
3	Jaar	Reclame-uitgaven (mio euro's)	Omzet (mio euro's)
4	20X0	0,51	5,0
5	20X1	0,42	4,5
6	20X2	0,33	4,0
7	20X3	0,45	3,5
8	20X4	0,61	6,0
9	20X5	0,56	5,5
10	20X6	0,22	3,8
11	20X7	0,39	4,8
12	20X8	0,58	6,1
13	20X9	0,70	6,4

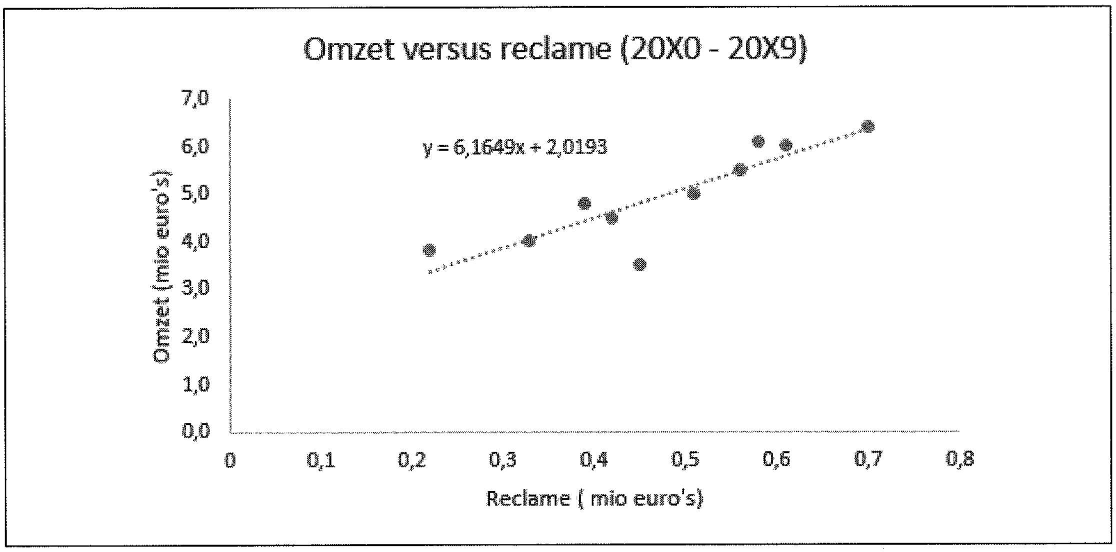
Eerst maken we een spreidingsdiagram, een grafiek van punten die bij elkaar behorende observaties van personen of objecten voorstellen. In ons geval is het object het jaartal en de reclame-uitgaven en de omzet het paar observaties van ieder jaartal. Volgende stappen moeten we uitvoeren:

> selecteer het gegevensbereik B4:C13, klik op het tabblad 'Invoegen' en kies in de groep 'Grafieken' de keuzelijst 'Spreidings- (X, Y) of bellendiagram invoegen' en klik op het subtype 'Spreiding'. Het resultaat is een spreidingsdiagram van gekoppelde observaties;
> klik op het tabblad 'Ontwerpen', dan op 'Grafiekonderdelen toevoegen' in de groep 'Grafiekindelingen', selecteer 'Astitels' en kies de optie 'Primair horizontaal'. Voer in het vak 'Astitel' de titel 'Reclame (mio euro's)' in. Selecteer nadien 'Astitels' opnieuw en klik op 'Primair verticaal'. Voer in het vak 'Astitel', 'Omzet (mio euro's)' in. Selecteer vervolgens nog de keuzelijst 'Grafiekonderdeel toevoegen' en klik op het onderdeel 'Grafiektitel'. Selecteer 'Boven grafiek' en vul in het label 'Grafiektitel', 'Omzet versus reclame (20X0-20X9)' in;
> verwijder ten slotte de rasterlijnen. Selecteer in de keuzelijst 'Grafiekonderdeel toevoegen' de optie 'Rasterlijnen' en klik vervolgens op de optie 'Primair groot verticaal' en nadien op de optie 'Primair groot horizontaal'. Het gewenste spreidingsdiagram is weergegeven in Figuur 11.



Figuur 11 Het spreidingsdiagram

In de bovenstaande grafiek kunnen we ook een trendlijn toevoegen. Daartoe klikken we met de rechtermuisknop op een punt in het spreidingsdiagram en er verschijnt een lijst waarin we 'Trendlijn toevoegen' selecteren. In het dialoogvenster 'Trendlijn opmaken' klikken we op het type 'Lineair' en vinken we ook de optie 'Vergelijking in grafiek weergeven' aan. In Figuur 12 zien we het resultaat.



Figuur 12 Spreidingsdiagram, regressielijn en regressievergelijking

Op de betekenis van de vergelijking gaan we later in.

hoofdstuk 4

Centrummaten

1 Het gemiddelde

1.1 Berekening

Het gemiddelde is een typische waarde voor een verzameling van waarden. De berekening gebeurt door de som van de waarden te delen door het aantal waarden.
Als n het aantal waarnemingen voorstelt, x een waarneming aanduidt, Σ staat voor 'tel alle waarnemingen op' en $\bar{x}$ het gemiddelde voorstelt, dan is de formule van het steekproefgemiddelde:

$$\bar{x} = \frac{1}{n} \Sigma x$$

Het gemiddelde μ van een populatie is:

$$\mu = \Sigma x \cdot p(x)$$

waarbij p(x) de kans is van een waarneming x in een populatie.

μ kan ook als volgt worden berekend:

$$\mu = \frac{\Sigma x}{N}$$

waarbij N dan staat voor het aantal waarnemingen in de populatie.

1.2 GEMIDDELDE (AVERAGE)

De statistische functie GEMIDDELDE berekent het rekenkundige gemiddelde van de getallen in een verzameling. De werkwijze is als volgt:

> je tikt een aantal cijfers in de cellen B1 tot B8 en selecteert de cel B9 voor het GEMIDDELDE;
> je klikt op de f_x-knop en opent het dialoogvenster 'Functie invoegen';
> je selecteert in de categorie 'Statistisch' de functie 'GEMIDDELDE' en het dialoogvenster 'Functieargumenten' verschijnt. In het venster bevat het vak 'Getal 1' al het bereik B1:B8 en wordt het gemiddelde gelijk aan 38,75 getoond;
> je klikt op 'OK', het venster 'Functieargumenten' verdwijnt en in het werkblad staat in cel B9 het resultaat.

Figuur 13 toont de cijfers en hun gemiddelde.

B9 | =GEMIDDELDE(B1:B8)

	A	B	C	D	E	F
1		38				
2		45				
3		36				
4		76				
5		24				
6		18				
7		9				
8		64				
9	GEMIDDELDE	38,75				
10						

Figuur 13 Berekening van het gemiddelde

1.3 GETRIMD. GEM (TRIMMEAN)

Deze functie elimineert extreme waarnemingen in een gegevensverzameling alvorens het gemiddelde te berekenen. Daardoor wordt het verkregen gemiddelde representatiever. De functieprocedure bestaat uit de volgende stappen:

> we tikken onze cijfergegevens in het bereik B1:B10 en selecteren B11 voor het resultaat;
> we klikken op f_x en openen het dialoogvenster 'Functie invoegen';
> we selecteren de statistische functie 'GETRIMD. GEM', klikken op 'OK' en het dialoogvenster 'Functieargumenten' verschijnt. In het Matrixvak voeren we B1:B10 in;
> in het Percentagevak vullen we het percentage scores dat we willen elimineren in als een cijfer achter de komma. We zullen 40 percent als extreme waarden elimineren en tikken 0,4 in. Dat betekent dat de 20 % grootste getallen en de 20 % kleinste getallen worden uitgesloten;
> we sluiten het dialoogvenster 'Functieargumenten' en in Figuur 14 lezen we dat ons gemiddelde gelijk is aan 648,33.

B11 | =GETRIMD.GEM(B1:B10;0,4)

	A	B	C	D	E	F
1		710				
2		600				
3		1290				
4		730				
5		675				
6		525				
7		1400				
8		250				
9		650				
10		180				
11	GETRIMD. GEM	648,33				
12						

Figuur 14 Berekening van het getrimd gemiddelde

2 Meetkundig gemiddelde

2.1 Berekening

Het meetkundig gemiddelde is een gemiddelde dat we berekenen als we rekening moeten houden met het multiplicatie-effect dat optreedt bij samengestelde intrestberekeningen, berekeningen van het gemiddelde inflatiepercentage of het gemiddelde groeipercentage ...
De formule voor het meetkundig gemiddelde M is:

$$M = \sqrt[n]{X_1 \times X_2 \ldots \times X_n}$$

2.2 MEETK. GEM (GEOMEAN)

De functie berekent het meetkundig gemiddelde van positieve numerieke gegevens.

Het gebruik wordt stapsgewijs getoond:

> je tikt een vijftal beursindexcijfers in en de bijbehorende jaarlijkse groeifactoren in het bereik B2:C6. In de cel B7 berekenen we het resultaat;
> je klikt op de f_x-knop en opent het dialoogvenster 'Functie invoegen' en selecteert 'MEETK. GEM';
> in het dialoogvenster 'Functieargumenten' tik je in het vak 'Getal 1' het bereik C3:C6 en het resultaat verschijnt: 1,062 afgerond;
> je klikt op 'OK', het venster 'Functieargumenten' wordt gesloten en in het werkblad lees je in cel B7 het meetkundig gemiddelde af.

Figuur 15 toont de ingevoerde gegevens en de waarde van het meetkundig gemiddelde.

B7 | f_x =MEETK.GEM(C3:C6)

	A	B	C	D	E
1	Jaar	Beursindexcijfer	Groeifactor		
2	1	100			
3	2	110	1,1		
4	3	112	1,02		
5	4	117,6	1,05		
6	5	127	1,08		
7	MEETK. GEM	1,062			
8					

Figuur 15 Berekening van het meetkundig gemiddelde

3 Harmonisch gemiddelde

3.1 Berekening

Het harmonisch gemiddelde H van een verzameling van N getallen x_1, x_2, x_3, …, x_N is het omgekeerde van het rekenkundig gemiddelde van de omgekeerden van die getallen. De formule voor het harmonisch gemiddelde H is:

$$H = \frac{N}{\sum \frac{1}{x}}$$

Het harmonisch gemiddelde wordt toegepast als het gemiddelde van samengestelde grootheden, zoals bijvoorbeeld de snelheid, berekend moet worden.

3.2 HARM. GEM (HARM. MEAN)

De functie berekent het harmonisch gemiddelde van N getallen in een verzameling. De functie werkt als volgt:

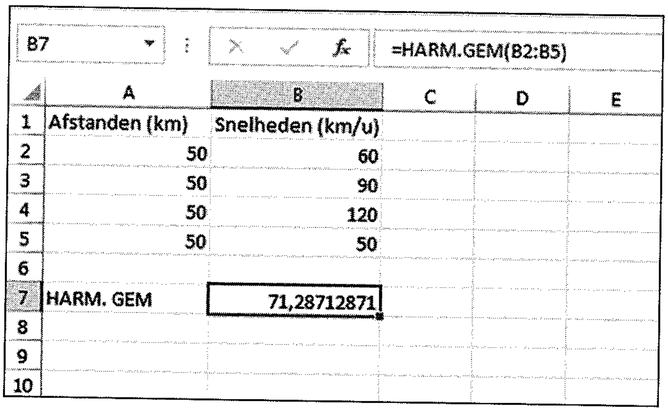
> je tikt een viertal snelheden in die tijdens verschillende gelijke stukken over een afstand van 200 km zijn gereden in het bereik van A2:B5. In de cel B7 berekenen we het resultaat of de gemiddelde snelheid;

> je klikt op de f_x-knop en opent het dialoogvenster 'Functie invoegen' en selecteert 'HARM. GEM';

> in het dialoogvenster 'Functieargumenten' tik je in het vak 'Getal 1' het bereik B2:B5 in en het resultaat verschijnt: 71,29 afgerond;

> je klikt op 'OK', het venster 'Functieargumenten' wordt gesloten en in het werkblad lees je in cel B7 het harmonisch gemiddelde af: 71,29 km/u.

Figuur 16 toont de ingevoerde gegevens en de waarde van het harmonisch gemiddelde.


B7 =HARM.GEM(B2:B5)

	A	B	C	D	E
1	Afstanden (km)	Snelheden (km/u)			
2	50	60			
3	50	90			
4	50	120			
5	50	50			
6					
7	HARM. GEM	71,28712871			
8					
9					
10					

Figuur 16 Berekening van het harmonisch gemiddelde

4 Mediaan

4.1 Berekening

De mediaan is de middelste waarde van een verzameling geordende getallen. Om bijvoorbeeld de mediaan te vinden van de maandelijkse geproduceerde hoeveelheden gsm-toestellen bij de nv MOBIEL in 20X1, ordenen we de cijfers in stijgende volgorde:

48 000, 49 000, 55 000, 56 000, 60 000, 62 000, 62 500, 62 700, 63 000, 63 200, 63 400, 63 500.

Omdat het aantal waarden even is, is de mediaan gelijk aan het gemiddelde van de twee middelste waarden of dus 62.250.

4.2 MEDIAAN (MEDIAN)

De functie MEDIAAN berekent de middelste waarde van een verzameling al of niet geordende gegevens. De volgende stappen zijn nodig:

> we voeren de hoeveelheden in het bereik B2:B13 in. Verder selecteren we B14 voor het resultaat;

> we klikken op de f_x-knop, openen het dialoogvenster 'Functie invoegen', selecteren 'MEDIAAN' en klikken op 'OK';

> in het dialoogvenster 'Functieargumenten' verschijnt de gegevensmatrix B2:B13 in het vak 'Getal 1' en onderaan de mediaanwaarde 62.250;

> we klikken op 'OK' in het venster 'Functieargumenten' en in het werkblad toont cel B14 het resultaat.

Figuur 17 toont het celbereik en de mediaanwaarde.

B14 | =MEDIAAN(B2:B13)

	A	B	C
1		Maandelijkse productie in 20X1 gerangschikt in stijgende volgorde	
2		48000	
3		49000	
4		55000	
5		56000	
6		60000	
7		62000	
8		62500	
9		62700	
10		63000	
11		63200	
12		63400	
13		63500	
14	MEDIAAN	62250	
15			

Figuur 17 Berekening van de mediaan

5 Modus

5.1 Berekening

De modus van een verzameling getallen of waarnemingen is het getal of de waarneming met de hoogste frequentie. Komt deze hoogste frequentie twee of meer keren voor, dan spreekt men van een bimodale of multimodale verzameling.

5.2 MODUS. ENKELV (MODE. SNGL)

De MODUS. ENKELV-functie werkt als volgt:

> we voeren de volgende getallenverzameling in:
50, 30, 66, 50, 69, 72, 69, 30, 45 en 69;
> in B12 berekenen we het modusresultaat;
> we klikken op de f_x-knop, het dialoogvenster 'Functie invoegen' verschijnt, we selecteren 'MODUS. ENKELV' en klikken op 'OK';
> in het dialoogvenster 'Functieargumenten' lezen we in het vak 'Getal 1' de gegevensmatrix B2:B11 en onderaan het resultaat, 69;
> we sluiten het dialoogvenster 'Functieargumenten' en in cel B12 van het werkblad verschijnt de modusuitkomst.

Figuur 18a toont het celbereik en de moduswaarde.

B12 =MODUS.ENKELV(B2:B11)

	A	B	C	D	E	F	G
1		Getallen					
2		50					
3		30					
4		66					
5		50					
6		69					
7		72					
8		69					
9		30					
10		45					
11		69					
12	MODUS	69					
13							

Figuur 18a Berekening van de modus

5.3 MODUS. MEERV (MODE. MULT)

Deze functie wordt gebruikt als er meerdere waarden in een gegevensverzameling de hoogste frequentie hebben.
De functie bestaat uit volgende stappen:

> we voeren volgend gegevensbereik in: 50, 30, 66, 50, 69, 72, 70, 30, 45, 69, 50, 41, 69;
> in B16:B17 berekenen we het resultaat;
> we klikken op de f_x-knop, het dialoogvenster 'Functie invoegen' verschijnt, we selecteren 'MODUS. MEERV' en klikken op 'OK';
> in het dialoogvenster 'Functieargumenten' voeren we in het vak 'Getal 1' de gegevensmatrix B2:B14 in;
> de functie berekent een verticale matrix van de modi;
> we drukken vervolgens op Ctrl + Shift + OK en in het werkblad verschijnen in de matrix B16:B17 de modiwaarden van het gegevensbereik.

Figuur 18b toont het celbereik en de modi.

B16 | =MODUS.MEERV(B2:B14)

	A	B	C	D	E	F	G
1		Getallen					
2		50					
3		30					
4		66					
5		50					
6		69					
7		72					
8		70					
9		30					
10		45					
11		69					
12		50					
13		41					
14		69					
15							
16	Modi	50					
17		69					
18							

Figuur 18b Berekening van de modi

hoofdstuk 5

Spreidingsmaten

1 Variatie

Variatie kunnen we omschrijven als de som van de gekwadrateerde afwijkingen van de waarnemingen uit een verzameling ten opzichte van hun groepsgemiddelde. Hoe groter deze afwijking is, hoe kleiner uiteraard de representativiteit van het gemiddelde is.

2 Populatievariantie

2.1 Berekening

Stel dat we het gewicht meten van 10 volwassenen en dat we de volgende uitkomsten vinden (in kg):

68, 72, 95, 87, 70, 99, 66, 91, 103, 79.

Als we het gemiddelde gewicht berekenen, dan vinden we 83 kg.
Tabel 4 toont ons de uitkomsten van onze waarnemingen en hun afwijkingen.

Tabel 4 Gewichten en hun afwijkingen

Gewicht	Afwijking ($x - \mu$)
68	-15
72	-11
95	+12
87	+4
70	-13
99	+16
66	-17
91	+8
103	+20
79	-4

Omdat het gemiddelde van de afwijkingen t.o.v. μ gelijk is aan nul, zullen we – om dit te voorkomen – de afwijkingen kwadrateren. Dat gemiddelde van de gekwadrateerde afwijkingen duiden we aan met de term variantie.
Tabel 5 toont nu Tabel 4 aangevuld met een derde kolom, zijnde de gekwadrateerde afwijkingen.

Tabel 5 Gewichten en hun gekwadrateerde afwijkingen

Gewicht	Afwijking $(x - \mu)$	Gekwadrateerde afwijking $(x - \mu)^2$
68	-15	225
72	-11	121
95	+12	144
87	+4	16
70	-13	169
99	+16	256
66	-17	289
91	+8	64
103	+20	400
79	-4	16

Als N het aantal gekwadrateerde afwijkingen voorstelt, dan is de formule voor de variantie aangeduid met het symbool σ^2:

$$\sigma^2 = \frac{\Sigma (x - \mu)^2}{N}$$

Toegepast op ons voorbeeld geeft dat:

$$\frac{225 + 121 + 144 + 16 + 169 + 256 + 289 + 64 + 400 + 16}{10} = 170$$

2.2 VAR. P (VAR. P)

De VAR. P-functie berekent de populatievariantie. De functie wordt als volgt gebruikt:

> in het werkblad zet je de gewichten uit Tabel 4 in het bereik B2:B11. Voor het resultaat kies je cel B12;

> je klikt op de f_x-knop om het dialoogvenster 'Functie invoegen' te openen. Je selecteert 'VAR. P' en klikt op 'OK';

> in het dialoogvenster 'Functieargumenten' voer je in het vak 'Getal 1' het bereik B2:B11 in en onderaan in het venster lezen we dat de populatievariantie gelijk is aan 170;

> je sluit het venster 'Functieargumenten' en in cel B12 van het werkblad staat de uitkomst van de populatievariantie σ^2.

Figuur 19 toont het werkblad met het celbereik en het resultaat.

B12 | f_x | =VAR.P(B2:B11)

	A	B	C	D	E
1		Gewichten (in kg)			
2		68			
3		72			
4		95			
5		87			
6		70			
7		99			
8		66			
9		91			
10		103			
11		79			
12	VAR.P	170			
13					

Figuur 19 Berekening van de populatievariantie

3 Steekproefvariantie

3.1 Berekening

De formule voor de berekening van de steekproefvariantie s^2 is:

$$s^2 = \frac{\Sigma (x - \bar{x})^2}{n}$$

Maar als we s^2 gebruiken om σ^2 te schatten, dan moeten we in de noemer $n - 1$ gebruiken en wordt de formule:

$$s^2 = \frac{\Sigma (x - \bar{x})^2}{n - 1}$$

Als we de gewichten uit Tabel 4 (p. 45) beschouwen als een steekproef getrokken uit een populatie met de bedoeling s^2 te berekenen als een schatter van σ^2, dan is s^2 gegeven de gekwadrateerde afwijkingen in Tabel 5 op p. 46 gelijk aan:

$$\frac{225 + 121 + 144 + 16 + 169 + 256 + 289 + 64 + 400 + 16}{10 - 1} = 188{,}88$$

3.2 VAR. S (VAR. S)

De VAR-functie berekent de steekproefvariantie. We herhalen dezelfde stappen als bij de VAR. P-functie:

> in het werkblad vul je de gewichten uit Tabel 4 in het bereik B2:B11 in en voor het resultaat selecteer je cel B12;
> je klikt op de f_x-knop, het dialoogvenster 'Functie invoegen' verschijnt. Je selecteert 'VAR. S', klikt op 'OK' en in het dialoogvenster 'Functieargumenten' is in het vak 'Getal 1' het bereik B2:B11 al ingevuld. Onderaan het venster lezen we het resultaat 188,88;
> je sluit het venster 'Functieargumenten' en het werkblad verschijnt met s^2 in cel B12.

Figuur 20 toont ons de berekening van s^2.

B12 =VAR.S(B2:B11)

	A	B	C	D
1		Gewichten (in kg)		
2		68		
3		72		
4		95		
5		87		
6		70		
7		99		
8		66		
9		91		
10		103		
11		79		
12	VAR.S	188,889		
13				

Figuur 20 Berekening van de steekproefvariantie

4 Populatiestandaarddeviatie

4.1 Berekening

Om de afwijking in zijn originele eenheid te kunnen uitdrukken, trekken we de vierkantswortel uit de variantie. Het resultaat noemen we de standaardafwijking σ. De formule is:

$$\sigma = \sqrt{\frac{\Sigma\,(x - \mu)^2}{N}}$$

Toegepast op ons voorbeeld vinden we:

$$\begin{aligned}\sigma &= \sqrt{170}\\ &= 13{,}04\end{aligned}$$

De populatiestandaarddeviatie is 13,04 kg.

4.2 STDEV. P (STDEV. P)

De Excelfunctie STDEV. P berekent de populatiestandaarddeviatie.
We doen het volgende:

> we voeren de gegevens uit Tabel 4 (p. 45) in het bereik B2:B11 in, en selecteren B12 voor het resultaat;
> we klikken op de f_x-knop en openen het dialoogvenster 'Functie invoegen'. Dan selecteren we 'STDEV. P' en klikken op 'OK' om het dialoogvenster 'Functieargumenten' te openen;
> in het dialoogvenster 'Functieargumenten' vinden we in vak 'Getal 1' het bereik B2:B11 en onderaan de populatiestandaarddeviatie;
> we sluiten het dialoogvenster 'Functieargumenten' en in cel B12 lezen we het resultaat, zijnde 13,038 kg.

Figuur 21 toont het werkblad met het celbereik en het resultaat voor σ.

B12 =STDEV.P(B2:B11)

	A	B	C	D
1		Gewichten (in kg)		
2		68		
3		72		
4		95		
5		87		
6		70		
7		99		
8		66		
9		91		
10		103		
11		79		
12	STDEV.P	13,038		
13				

Figuur 21 Berekening van de populatiestandaarddeviatie

5 Steekproefstandaardafwijking

5.1 Berekening

De standaardafwijking van een steekproef, beschouwd als een schatter van de populatiestandaarddeviatie, wordt voorgesteld door het symbool s en is gelijk aan de vierkantswortel uit s^2 of:

$$s = \sqrt{\frac{\Sigma\,(x - \bar{x})^2}{n - 1}}$$

Met de VAR. S (VAR. S)-functie hadden we voor s^2 een waarde gevonden gelijk aan 188,88. Dat levert voor s op:

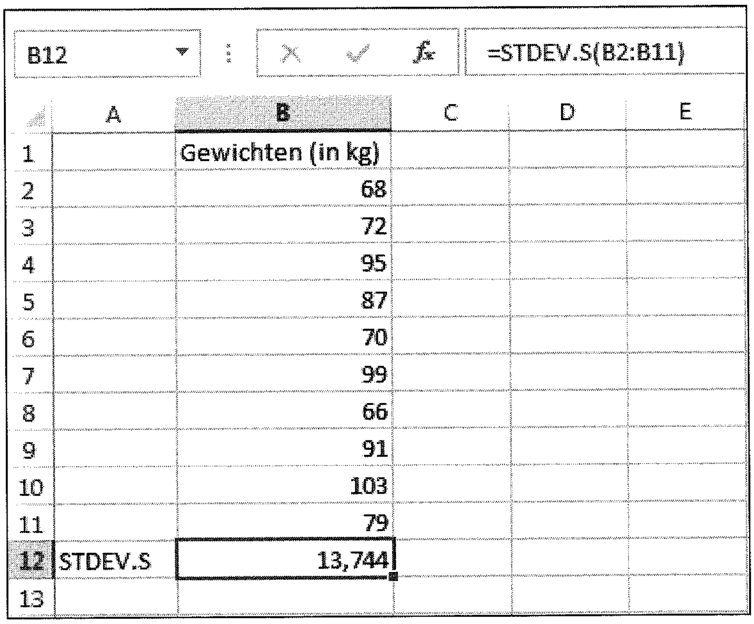
$$\begin{aligned} s &= \sqrt{188{,}88} \\ &= 13{,}74 \end{aligned}$$

De steekproefstandaarddeviatie is gelijk aan 13,74 kg.

5.2 STDEV. S (STDEV. S)

De functie STDEV. S berekent de steekproefstandaardafwijking.
Hoe gebruik je deze functie?

- In het werkblad voer je de gegevens in het bereik B2:B11 terug in en selecteer je B12 voor het resultaat.
- Je klikt op de f_x-knop en opent het venster Functie invoegen, je selecteert 'STDEV. S', klikt op 'OK' en opent het dialoogvenster 'Functieargumenten'.
- In het venster 'Functieargumenten' is het bereik B2:B11 al in het vak 'Getal 1' geplaatst en lezen we onderaan het resultaat af, 13,744 kg.
- Je sluit het dialoogvenster en in Figuur 22 vinden we in cel B12 de steekproefstandaardafwijking voor het bereik B2:B11.


B12 =STDEV.S(B2:B11)

	A	B	C	D	E
1		Gewichten (in kg)			
2		68			
3		72			
4		95			
5		87			
6		70			
7		99			
8		66			
9		91			
10		103			
11		79			
12	STDEV.S	13,744			
13					

Figuur 22 Berekening van de steekproefstandaardafwijking

hoofdstuk 6

Rangschikken

1 Standaardiseren

1.1 Berekening

Deze techniek wordt gebruikt om diverse waarnemingen uit diverse verzamelingen met elkaar te vergelijken. Daartoe standaardiseren we eerst de geobserveerde waarden. De nieuwe waarden noemen we Z-waarden.

De formule voor de berekening van Z in geval van een populatie is:

$$Z = \frac{x - \mu}{\sigma}$$ [1]

Om de Z-waarde te vinden, delen we de afwijking van x ten opzichte van zijn gemiddelde μ door de standaarddeviatie σ van de populatie. Z kan positief, negatief of nul zijn en μ en σ van alle Z-waarden zijn respectievelijk 0 en 1.

Als we deze conversie voor meerdere verzamelingen van waarnemingen hebben gedaan, kunnen we de waarnemingen uit de verschillende verzamelingen met elkaar vergelijken.

1.2 NORMALISEREN (STANDARDIZE)

Examenuitslagen voor wiskunde betekenen bijvoorbeeld niet hetzelfde als examenuitslagen voor het vak geschiedenis. Om ze vergelijkbaar te maken, moeten we de uitslagen van ieder examen converteren in Z-scores.

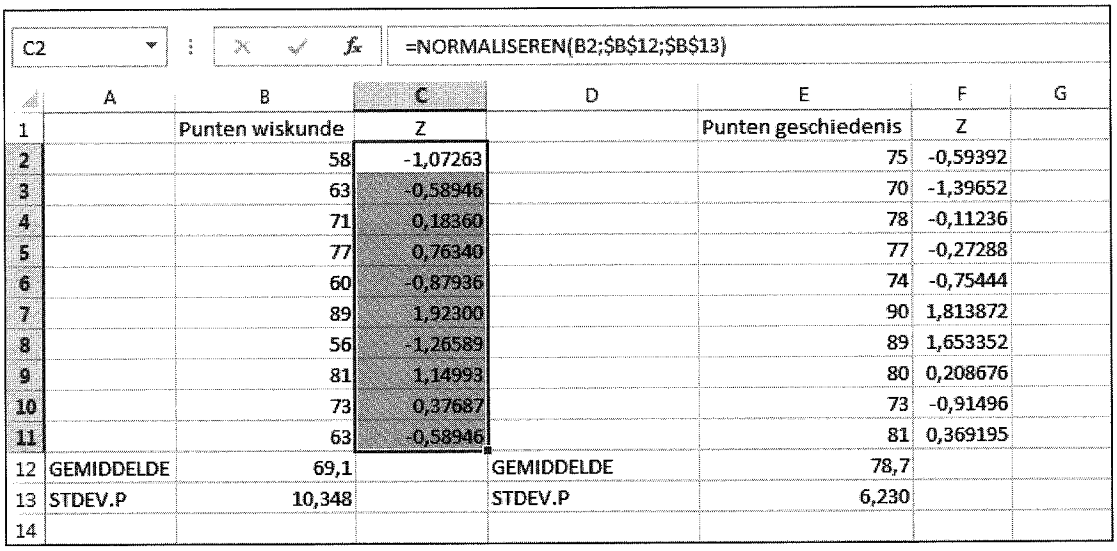
De werkbladfunctie NORMALISEREN berekent Z-scores. In figuur 23 tonen we de resultaten van studenten voor wiskunde en geschiedenis. We berekenen voor beide verzamelingen het gemiddelde en de standaarddeviatie en maken daarbij gebruik van de functies GEMIDDELDE en STDEV. P, die we eerder al behandeld hebben. We gaan nu na hoe we de punten voor het vak wiskunde normaliseren:

> we selecteren in het werkblad de cel C2 en klikken op de f_x-knop om het venster 'Functie invoegen' te openen;
> in het venster 'Functie invoegen' selecteren we de functie 'NORMALISEREN', klikken op 'OK' en het venster 'Functieargumenten' verschijnt;
> in het X-vak tikken we de celcoördinaten in van het eerste resultaat voor het vak wiskunde of B2;
> in het vak 'Gemiddelde' voeren we de cel in van het gemiddelde voor wiskunde, B12. We maken B12 absoluut door op de toets F4 te drukken;

[1] In geval van een steekproef is $Z = \frac{x - \bar{x}}{s}$.

> in het vak 'Standaarddev' brengen we de coördinaten in van de cel die de standaarddeviatie bevat, B13. Ook hier maken we B13 absoluut door F4 in te drukken;
> we klikken op 'OK' en de Z-waarde voor de eerste examenuitslag verschijnt in cel C2;
> vervolgens klikken we in de rechterbenedenhoek van cel C2, plaatsen de muisaanwijzer op de vulgreep, de vorm van de muisaanwijzer verandert in een zwart kruis. We voeren de functie NORMALISEREN door tot en met de cel C11.

We gaan op eenzelfde manier te werk voor het normaliseren van de punten voor het vak geschiedenis. Figuur 23 toont ons de Z-waarden voor alle punten van wiskunde en geschiedenis.



C2 =NORMALISEREN(B2;$B$12;$B$13)

	A	B	C	D	E	F	G
1		Punten wiskunde	Z		Punten geschiedenis	Z	
2		58	-1,07263		75	-0,59392	
3		63	-0,58946		70	-1,39652	
4		71	0,18360		78	-0,11236	
5		77	0,76340		77	-0,27288	
6		60	-0,87936		74	-0,75444	
7		89	1,92300		90	1,813872	
8		56	-1,26589		89	1,653352	
9		81	1,14993		80	0,208676	
10		73	0,37687		73	-0,91496	
11		63	-0,58946		81	0,369195	
12	GEMIDDELDE	69,1		GEMIDDELDE	78,7		
13	STDEV.P	10,348		STDEV.P	6,230		
14							

Figuur 23 Berekening van Z-waarden

Als we nu de Z-waarde van een uitslag voor wiskunde van 73 punten vergelijken met de Z-waarde van diezelfde uitslag voor het vak geschiedenis, dan stellen we vast dat de Z-waarden verschillen: 0,376869 en -0,914962. Het is nu duidelijk dat het resultaat van 73 punten voor wiskunde beter is dan de 73 punten behaald voor geschiedenis, aangezien de Z-waarde voor wiskunde groter is dan de Z-waarde voor geschiedenis.

2 Rangschikken

2.1 RANG. GELIJK (RANK. EQ)

Met de rang van bijvoorbeeld een examenuitslag bedoelen we de positie van die examenuitslag binnen een verzameling van examenuitslagen. Het is evident dat de hoogste uitslag rang 1 zal hebben, de tweede uitslag rang 2 enz.

Met de functie RANG. GELIJK kun je de posities van bijvoorbeeld alle examenuitslagen voor een vak vlug bepalen. Dit is de te volgen procedure:

> je voert de gegevens in het bereik B2:B11 in en selecteert cel C2;
> je klikt op de f_x-knop en het venster 'Functie invoegen' wordt geopend;
> nu selecteer je de functie 'RANG. GELIJK' en opent het venster 'Functieargumenten';
> in het vak 'Getal' voer je de cel in van de eerste uitslag waarvan we de rang willen invoeren in cel C2;
> in het vak 'Verw' voer je de kolom B2:B11 in en maakt deze absoluut door op de F4-toets te drukken;
> in het vak 'Volgorde' vul je niets in of het getal 0. De uitslagen worden nu in dalende volgorde gerangschikt. Als je een waarde ongelijk aan nul invoert, dan worden de uitslagen in oplopende volgorde gerangschikt;
> klik nu op 'OK' om de rang in cel C2 in te voeren;
> je klikt in de rechterbenedenhoek van cel C2, plaatst de muisaanwijzer op de vulgreep, de vorm van de muisaanwijzer verandert in een zwart kruis en je voert de functie 'RANG. GELIJK' door tot en met de cel C11. Het resultaat wordt getoond in Figuur 24.

C2 =RANG.GELIJK(B2;$B$2:$B$11;0)

	A	B	C	D	E	F	G
1		Punten	Rang				
2		68	9				
3		72	7				
4		95	2				
5		87	4				
6		70	8				
7		99	1				
8		66	10				
9		91	3				
10		83	5				
11		79	6				
12							

Figuur 24 Berekening van de positie van een examenuitslag

2.2 GROOTSTE (LARGE) en KLEINSTE (SMALL)

We kunnen het proces van de rangbepaling ook omkeren en ons afvragen welke uitslag overeenstemt met een welbepaalde rang. De functies GROOTSTE en KLEINSTE berekenen respectievelijk de op K – 1 na grootste waarde en de op K – 1 na kleinste waarde in een verzameling.

Figuur 25a bevat een werkblad met examenuitslagen voor het vak Frans. We willen nu de vierde grootste uitslag vinden en selecteren daarom in het venster Functie invoegen de functie GROOTSTE. Vervolgens openen we het venster Functieargumenten en vullen in het vak Matrix het celbereik B2:B11 in. In het vak K zetten we rang 4 waarvan we dus de corresponderende uitslag willen vinden. We klikken op 'OK' en in de geselecteerde cel B12 verschijnt het resultaat 73.
Willen we bijvoorbeeld de vijfde kleinste score kennen, dan selecteren we in het venster Functie invoegen de functie KLEINSTE. We openen terug het venster Functieargumenten en vullen in het vak Matrix het celbereik B2:B11 in. In het vak K vullen we rang 5 in, waarvan we de overeenstemmende uitslag willen weten. We klikken op 'OK' en in de geselecteerde cel B12 van figuur 25b krijgen we het resultaat 63 als vijfde laagste score.

B12 | fx =GROOTSTE(B2:B11;4)

	A	B	C	D	E
1		Punten Frans			
2		58			
3		63			
4		71			
5		77			
6		60			
7		89			
8		56			
9		81			
10		73			
11		62			
12	GROOTSTE	73			
13					

Figuur 25a Berekening van de vierde grootste score

B12 | fx =KLEINSTE(B2:B11;5)

	A	B	C	D	E
1		Punten Frans			
2		58			
3		63			
4		71			
5		77			
6		60			
7		89			
8		56			
9		81			
10		73			
11		62			
12	KLEINSTE	63			
13					

Figuur 25b Berekening van de vijfde kleinste score

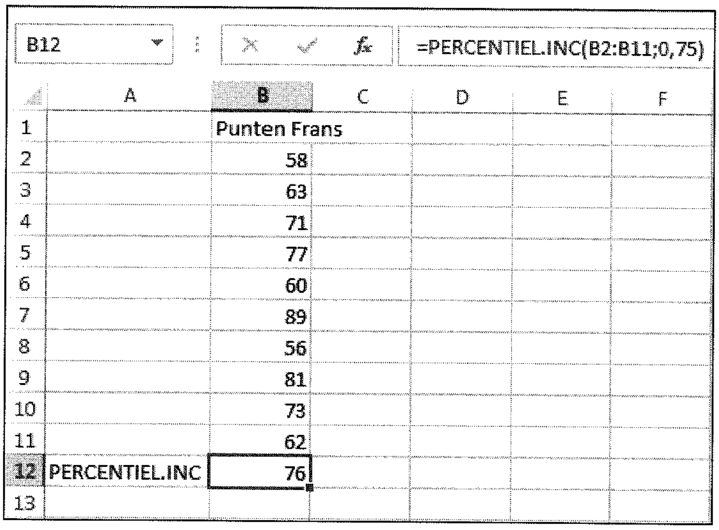
2.3 PERCENTIEL. INC (PERCENTILE. INC)

Het percentiel stelt een relatieve positie voor van een waarneming in een verzameling. Als een waarneming overeenkomt met bijvoorbeeld het tachtigste percentiel, betekent dit dat 80 % van de waarnemingen daaraan gelijk is of kleiner is.

De PERCENTIEL. INC-functie berekent alle mogelijke percentielen uit een verzameling van waarnemingen. Hierna geven we aan hoe de functie wordt gebruikt:

> we voeren in het bereik B2:B11 de resultaten in van een examen Frans en selecteren cel B12;
> we klikken op de f_x-knop om het venster 'Functie invoegen' te openen;
> we selecteren de functie 'PERCENTIEL. INC', klikken op 'OK' en openen het venster 'Functieargumenten';
> in het vak 'Matrix' voeren we het bereik B2:B11 in;
> in het vak 'K' vullen we het te berekenen percentiel in als een decimaal getal. Om het 75ste percentiel te vinden, voeren we dus 0,75 in;
> we klikken op 'OK' en in cel B12 verschijnt de waarde 76. Dit betekent dat 75 % van de uitslagen kleiner of gelijk is aan 76.

Figuur 26 toont het resultaat in het werkblad.



B12 =PERCENTIEL.INC(B2:B11;0,75)

	A	B	C	D	E	F
1		Punten Frans				
2		58				
3		63				
4		71				
5		77				
6		60				
7		89				
8		56				
9		81				
10		73				
11		62				
12	PERCENTIEL.INC	76				
13						

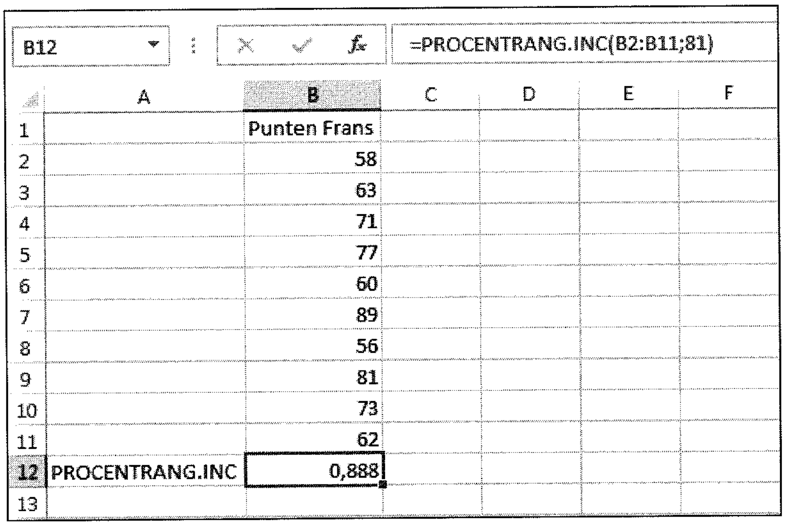
Figuur 26 Berekening van het 75ste percentiel

2.4 PROCENTRANG. INC (PERCENTRANK. INC)

We kunnen ook geïnteresseerd zijn in de relatieve positie van een waarneming in een verzameling van gegevens. Daarvoor kun je de Excelfunctie PROCENTRANG. INC gebruiken.

We hernemen de gegevens over het examen Frans en we willen nu de relatieve positie kennen van de uitslag 81:

> in het venster 'Functie invoegen', dat verschijnt na het indrukken van de f_x-knop, selecteren we de functie 'PROCENTRANG. INC';
> in het dialoogvenster 'Functieargumenten' voeren we in het vak Matrix' het bereik B2:B11 in; in het X-vak de waarde '81' waarvan we de positie willen kennen;
> onderaan verschijnt het resultaat 0,888 of 88,8 %;
> we klikken op 'OK' en in Figuur 27 verschijnt in de geselecteerde cel B12 in het werkblad ook het resultaat.


B12 =PROCENTRANG.INC(B2:B11;81)

	A	B	C	D	E	F
1		Punten Frans				
2		58				
3		63				
4		71				
5		77				
6		60				
7		89				
8		56				
9		81				
10		73				
11		62				
12	PROCENTRANG.INC	0,888				
13						

Figuur 27 Berekening van de positie van het cijfer 81

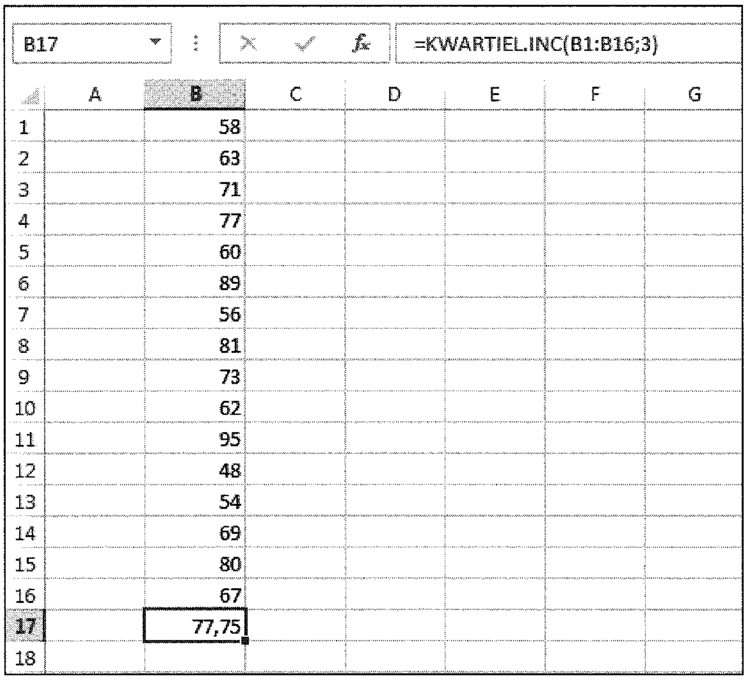
2.5 KWARTIEL. INC (QUARTILE. INC)

Specifieke percentielen van een verzameling gegevens zijn het 25ste, het 50ste, het 75ste en het 100ste percentiel. Zij verdelen de gegevens in kwarten en daarom worden ze kwartielen genoemd.

Met de functie KWARTIEL. INC kunnen we deze kwartielen berekenen. In Figuur 28 hebben we in het bereik B1:B16 een aantal scores ingevuld. Van deze verzameling kunnen we de vier kwartielen berekenen. Hieronder vind je de werkwijze voor de berekening van het derde kwartiel:

> met de f_x-knop open je het venster 'Functie invoegen' en selecteer je de functie KWARTIEL. INC. Klik op 'OK';
> het venster 'Functieargumenten' verschijnt. In het vak 'Matrix' voer je het bereik B1:B16 in waarvoor we de kwartielen kunnen berekenen;

> in het vak 'Kwartiel' vul je 0, 1, 2, 3 of 4 in. Deze getallen corresponderen respectievelijk met de minimumwaarde, het 1ste kwartiel, het 2de kwartiel, het 3de en het 4de kwartiel of de maximumwaarde van de verzameling.
  Als je 3 invoert, dan is het resultaat 77,75, m.a.w. 75 % van de scores is gelijk aan 77,75 of minder;
> klik in het venster op 'OK' en in de geselecteerde cel B17 verschijnt ook het 3de kwartiel.



B17 =KWARTIEL.INC(B1:B16;3)

	A	B	C	D	E	F	G
1		58					
2		63					
3		71					
4		77					
5		60					
6		89					
7		56					
8		81					
9		73					
10		62					
11		95					
12		48					
13		54					
14		69					
15		80					
16		67					
17		77,75					
18							

Figuur 28 Berekening van het 3de kwartiel

2.6 Rang en percentiel (Rank and percentile)

De toepassing 'Rang en percentiel' is een gegevensanalysetechniek die tegelijkertijd van ieder gegeven uit een verzameling de rang en het percentiel berekent.

In Figuur 29 hernemen we de gegevens die we voor het toepassen van de KWARTIEL.INC-functie gebruikt hebben. We gaan als volgt te werk:

> in het lint klikken we op het tabblad 'Gegevens' en selecteren de keuzelijst 'Gegevensanalyse';
> in het venster 'Gegevensanalyse' selecteren we de functie 'Rang en percentiel', klikken op 'OK' en het dialoogvenster 'Rang en percentiel' verschijnt;
> in het vak Invoerbereik voeren we het volledige gegevensbereik in: B1:B17. De matrix wordt automatisch absoluut gemaakt. Verder kiezen we de optie 'Groeperen per kolom', aangezien de gegevens in kolommen worden weergegeven;

> ten slotte selecteren we als uitvoeroptie de knop 'Nieuw werkblad' en klikken we op 'OK' om het venster 'Rang en percentiel' te sluiten.

In Figuur 30 zien we dat de uitslagen in de kolommen *Punten* en *Rang* gerangschikt zijn van hoog naar laag. De kolom *Gegevenspunt* duidt de positie van de uitslagen aan in het oorspronkelijke gegevensbereik. De kolom *Procent* geeft voor iedere uitslag het overeenstemmende percentiel.

	A	B	C
1		Punten	
2		58	
3		63	
4		71	
5		77	
6		60	
7		89	
8		56	
9		81	
10		73	
11		62	
12		95	
13		48	
14		54	
15		69	
16		80	
17		67	
18			

Figuur 29 Gegevensverzameling voor de toepassing van de functie Rang en percentiel

	A	B	C	D	E
1	*Gegevenspunt*	*Punten*	*Rang*	*Procent*	
2	11	95	1	100,00%	
3	6	89	2	93,30%	
4	8	81	3	86,60%	
5	15	80	4	80,00%	
6	4	77	5	73,30%	
7	9	73	6	66,60%	
8	3	71	7	60,00%	
9	14	69	8	53,30%	
10	16	67	9	46,60%	
11	2	63	10	40,00%	
12	10	62	11	33,30%	
13	5	60	12	26,60%	
14	1	58	13	20,00%	
15	7	56	14	13,30%	
16	13	54	15	6,60%	
17	12	48	16	0,00%	
18					

Figuur 30 Resultaat van de functie Rang en percentiel

hoofdstuk 7

Indexcijfers

1 Definitie

Een indexcijfer meet met hoeveel een variabele grootheid op een bepaald tijdstip verandert in verhouding tot haar waarde in een bepaalde basisperiode.
De berekening van een indexcijfer gebeurt als volgt:

$$\text{Indexcijfer} = \frac{\text{actuele waarde}}{\text{basiswaarde}} \times 100$$

Het indexcijfer is een percentage met weglating van het procentteken.

We geven het volgende voorbeeld. Veronderstel dat een producent van fietsen in het jaar 20x1 25 000 eenheden verkocht en dat het verkochte aantal in 20x2 opliep tot 27 250 eenheden. Dan is het indexcijfer:

$$\frac{27\,250}{25\,000} \times 100 = 109$$

Het indexcijfer meet de verandering van het aantal verkochte fietsen in 20x2 ten opzichte van de afzet in het basisjaar 20x1: + 9 %.
Merk op dat het indexcijfer voor de basisperiode altijd gelijk is aan 100.

2 Prijsindexcijfer

Een prijsindexcijfer meet de prijsverandering van een goed of van een korf van goederen op een bepaald tijdstip ten opzichte van een bepaalde basisperiode.

In het geval van consumenten meet het de verandering van de kosten van levensonderhoud.

2.1 Enkelvoudig prijsindexcijfer

Het enkelvoudig prijsindexcijfer berekent de prijsverandering van een goed op een bepaald tijdstip ten opzichte van een bepaalde basisperiode.

De formule voor het enkelvoudig prijsindexcijfer I_p^e is:

$$I_p^e = \frac{p_t}{p_0} \times 100$$

waarbij:
p_t = prijs van een goed in periode t
p_0 = prijs van een goed in de basisperiode

Voorbeeld
Een brood kostte in 20x1 € 1,25 en in 20x2 € 1,50. I_p^e voor 20x2 wordt:

$$I_p^e = \frac{1,50}{1,25} \times 100 = 120$$

De broodprijs is dus gestegen met 20 %, wat overeenkomt met het verschil tussen het indexcijfer van 20x2 en het indexcijfer van 20x1.

2.2 Samengesteld prijsindexcijfer

Het samengesteld prijsindexcijfer meet in één cijfer de prijsveranderingen van verschillende goederen op een bepaald tijdstip ten opzichte van een bepaalde basisperiode. In wat volgt bespreken we de mogelijkheden om dat indexcijfer te berekenen.

2.2.1 Ongewogen samengesteld prijsindexcijfer

We berekenen het ongewogen samengesteld prijsindexcijfer voor een korf van verschillende goederen door de som van de prijzen van die goederen op een bepaald tijdstip te delen door de som van de prijzen van die goederen in de basisperiode.
De formule voor dit samengesteld prijsindexcijfer I_p^s is:

$$I_p^s = \frac{\sum p_t}{\sum p_0} \times 100$$

Voorbeeld
In onderstaande tabel geven we de prijzen van drie voedingswaren voor het basisjaar 20x1 en het actuele jaar 20x2:

Tabel 6 Jaarlijkse prijzen van drie voedingswaren

Voedingswaar	20x1	20x2
Brood	€ 1,00	€ 1,20
Melk per liter	€ 0,80	€ 1,00
Boter per kg	€ 1,25	€ 1,60

$$I_p^s = \frac{1,20 + 1 + 1,60}{1 + 0,8 + 1,25} \times 100 = \frac{3,8}{3,05} \times 100 = 124,59$$

De prijzen van deze goederen stegen van 20x1 tot 20x2 met 24,59 %

Een tekortkoming van het ongewogen samengesteld prijsindexcijfer is dat de consumptie van deze producten als egaal wordt geïnterpreteerd, terwijl dat niet het geval is. Om een correcter samengesteld prijsindexcijfer te berekenen moeten we rekening houden met de ongelijke consumptie van deze goederen. Daarom gaan we de prijs van elk goed wegen met de corresponderende hoeveelheid die ervan verbruikt wordt.

2.2.2 Gewogen samengesteld prijsindexcijfer

Het gewogen samengesteld prijsindexcijfer wordt dus berekend door de prijzen van de verschillende goederen te wegen met de verbruikte hoeveelheden ofwel uit het basisjaar, ofwel uit de actuele periode.

A Laspeyres-prijsindexcijfer

Bij de berekening van dit prijsindexcijfer worden de hoeveelheden uit de basisperiode gebruikt om de prijzen van de goederen te wegen.
De formule is dan:

$$I_p^L = \frac{\sum p_t \times q_0}{\sum p_0 \times q_0} \times 100$$

waarbij:
q_0 = hoeveelheid van een goed in de basisperiode

Voorbeeld
In de volgende tabel geven we naast de prijzen van de producten in 20x1 en 20x2 ook het aantal consumpties gedurende de basisperiode 20x1 weer:

Tabel 7 Jaarlijkse prijzen van drie voedingswaren en hun verbruik in het basisjaar 20x1

Voedingswaar	Prijs in 20x1	Prijs in 20x2	Hoeveelheid in 20x1
Brood	€ 1,00	€ 1,20	6 000 000
Melk per liter	€ 0,80	€ 1,00	7 000 000
Boter per kg	€ 1,25	€ 1,60	2 000 000

We berekenen nu het Laspeyres-indexcijfer voor 20x2:

$$I_p^l = \frac{(6\,000\,000 \times 1,20) + (7\,000\,000 \times 1) + (2\,000\,000 \times 1,60)}{(6\,000\,000 \times 1) + (7\,000\,000 \times 0,80) + (2\,000\,000 \times 1,25)} \times 100 = 123,40$$

Uit de berekening blijkt dat de prijzen tussen 20x1 en 20x2 zijn gestegen met 23,40 %. Het berekend indexcijfer is lager dan het ongewogen samengesteld prijsindexcijfer omdat boter, waarvan de prijswijziging het grootst is, de kleinste weging heeft.

Het nadeel van deze methode is dat de historische hoeveelheden van de basisperiode veel kunnen afwijken van de hoeveelheden in de actuele periode. De historische hoeveelheden zijn in dat geval weinig representatief voor de actuele periode.

B Paasche-prijsindexcijfer

Voor de berekening van dit prijsindexcijfer worden de hoeveelheden uit de actuele periode als wegingsfactoren gebruikt.

De formule is dus:

$$I_p^P = \frac{\sum p_t \times q_t}{\sum p_0 \times q_t} \times 100$$

waarbij:
q_t = hoeveelheid van een goed in periode t

Voorbeeld
We hernemen het vorige voorbeeld maar vervangen de hoeveelheden van 20x1 door de hoeveelheden van 20x2:

Tabel 8 Jaarlijkse prijzen van drie voedingswaren en hun verbruik in het actuele jaar 20x2

Voedingswaar	Prijs in 20x1	Prijs in 20x2	Hoeveelheid in 20x2
Brood	€ 1,00	€ 1,20	5 700 000
Melk per liter	€ 0,80	€ 1,00	6 500 000
Boter per kg	€ 1,25	€ 1,60	1 800 000

Met deze gegevens berekenen we I_p^P voor het jaar 20x2:

$$I_p^P = \frac{(5\,700\,000 \times 1,20) + (6\,500\,000 \times 1) + (1\,800\,000 \times 1,60)}{(5\,700\,000 \times 1) + (6\,500\,000 + 0,80) + (1\,800\,000 \times 1,25)} \times 100 = 123,35$$

De prijzen stegen in 20x2 met 23,35 % ten opzichte van het basisjaar 20x1.

Het indexcijfer is kleiner dan dat van Laspeyres omdat de wegingen van melk en boter waarvan de prijswijzigingen het grootst zijn, nu kleiner zijn en de weging van brood, waarvan de prijswijziging het kleinst is, gestegen is.

Het Paasche-prijsindexcijfer is interessanter dan dat van Laspeyres omdat de berekening van I_p^P gebruikmaakt van actuele hoeveelheden.

C Fisher-prijsindexcijfer

Het prijsindexcijfer van Laspeyres is in het geval van prijsverhogingen in het algemeen te groot omdat bij Laspeyres de actuele prijzen in de teller worden gewogen met hoeveelheden uit de basisperiode en er dus geen rekening wordt gehouden met een daling van het verbruik.

In het geval van prijsverlagingen stellen we het omgekeerde vast, dan zal het indexcijfer te klein zijn vermits de gedaalde prijzen in de teller terug worden gewogen met hoeveelheden uit de basisperiode en er dus geen rekening wordt gehouden met een stijging van het verbruik.

Bij Paasche leiden prijsverhogingen ook tot een overschatting van het indexcijfer en prijsverlagingen eveneens tot een onderschatting van het indexcijfer.

De Amerikaanse econoom Irving Fisher ontwikkelde een gewogen samengesteld prijsindexcijfer dat rekening hield met bovengenoemde gebreken door de berekening van het meetkundig gemiddelde van het Laspeyres- en Paasche-prijsindexcijfer.

De formule van dit indexcijfer is :

$$I_p^F = \sqrt{\frac{\sum p_t \times q_0}{\sum p_0 \times q_0} \times 100 \times \frac{\sum p_t \times q_t}{\sum p_0 \times q_t} \times 100}$$

of

$$I_p^F = \sqrt{I_p^L \times I_p^P}$$

Op basis van de uitkomsten die hierboven gevonden werden voor I_p^L en I_p^P berekenen we het gewogen samengesteld prijsindexcijfer I_p^F voor het jaar 20x2:

$$I_p^F = \sqrt{123,40 \times 123,35} = 123,38$$

In 20x2 zijn de prijzen gestegen met 23,38 % tegenover de prijzen die van toepassing waren in 20x1. Het Fisher indexcijfer voorkomt in ons voorbeeld de onjuiste berekening van de prijsstijging opgeleverd door de prijs-indexcijfers van Laspeyres en Paasche.

3 Hoeveelheidsindexcijfer

Een hoeveelheidsindexcijfer meet de hoeveelheidsverandering van een goed of van een korf van goederen op een bepaald tijdstip ten opzichte van een bepaalde basisperiode.

Voor consumenten meet het indexcijfer dan bijvoorbeeld de verandering van de consumptie.

3.1 Enkelvoudig hoeveelheidsindexcijfer

Het enkelvoudig hoeveelheidsindexcijfer berekent de hoeveelheidsverandering van een goed op een bepaald tijdstip ten opzichte van een bepaalde periode.

De formule is:

$$I_q^e = \frac{q_t}{q_0} \times 100$$

waarbij:
q_t = hoeveelheid van een goed in periode t
q_0 = hoeveelheid van een goed in de basisperiode

Voorbeeld
Een bedrijf verkocht in 20x1 10 000 ijskasten en in 20x2 steeg de afzet tot 10 500 eenheden. Het I_q^e voor 20x2 is dan:

$$I_q^e = \frac{10\,500}{10\,000} \times 100 = 105$$

De afzet is in 20x2 gestegen met 5 %.

3.2 Samengesteld hoeveelheidsindexcijfer

Het samengesteld hoeveelheidsindexcijfer berekent in één cijfer de hoeveelheidsveranderingen van verschillende goederen op een bepaald tijdstip ten opzichte van een bepaalde basisperiode. Hierna bekijken we de verschillende samengestelde hoeveelheidsindexcijfers die berekend kunnen worden.

3.2.1 Ongewogen samengesteld hoeveelheidsindexcijfer

We berekenen het ongewogen samengesteld hoeveelheidsindexcijfer voor een korf van verschillende goederen door de som van de hoeveelheden van die goederen op een bepaald tijdstip te delen door de som van de hoeveelheden van die goederen in de basisperiode.

De formule voor dit indexcijfer is:

$$I_q^s = \frac{\sum q_t}{\sum q_0} \times 100$$

Voorbeeld

In de volgende tabel geven we de hoeveelheden van drie artikelen die door een elektronicabedrijf werden verkocht in 20x1 en 20x2:

Tabel 9 Jaarlijkse verkochte hoeveelheden van drie producten

Product	20x1	20x2
Radio	10 000	8 300
Dvd-speler	15 000	14 200
Tv	25 000	18 500

$$I_q^s = \frac{8\,300 + 14\,200 + 18\,500}{10\,000 + 15\,000 + 25\,000} \times 100 = \frac{41\,000}{50\,000} \times 100 = 82$$

De verkoop van deze goederen is in 20x2 met 18 % gedaald ten opzichte van het basisjaar 20x1.

Bij de berekening van dit indexcijfer werd verondersteld dat de verkoopprijzen van de drie producten gelijk zijn. Maar dat is niet zo. Om een juister hoeveelheidsindexcijfer te berekenen, moeten we de hoeveelheden van elk goed wegen met hun overeenstemmende verkoopprijs.

3.2.2 Gewogen samengesteld hoeveelheidsindexcijfer

Het gewogen samengesteld hoeveelheidsindexcijfer berekenen we door de hoeveelheden van de verschillende goederen te wegen met de verkoopprijzen hetzij uit het basisjaar ofwel uit de actuele periode.

A Laspeyres-hoeveelheidsindexcijfer

Bij de berekening van het Laspeyres-hoeveelheidsindexcijfer worden de prijzen uit de basisperiode gebruikt om de hoeveelheden van de goederen te wegen.

De formule hiervoor is:

$$I_q^L = \frac{\sum q_t \times p_0}{\sum q_0 \times p_0} \times 100$$

Voorbeeld

In de volgende tabel geven we naast de verkochte hoeveelheden van de producten in 20x1 en 20x2 ook de verkoopprijzen die golden in basisperiode 20x1 weer:

Tabel 10 Jaarlijkse verkochte hoeveelheden van drie producten en hun prijzen in het basisjaar 20x1

Product	Hoeveelheid in 20x1	Hoeveelheid in 20x2	Prijs in 20x1
Radio	10 000	8 300	€ 50
Dvd-speler	15 000	14 200	€ 100
Tv	25 000	18 500	€ 500

Het Laspeyres-indexcijfer voor 20x2 is dan:

$$I_q^L = \frac{(8\,300 \times 50) + (14\,200 \times 100) + (18\,500 \times 500)}{(10\,000 \times 50) + (15\,000 \times 100) + (25\,000 \times 500)} \times 100 = \frac{11\,085\,000}{1\,450\,000} \times 100 = 76{,}45$$

De verkochte hoeveelheid in 20x2 is ten opzichte van het basisjaar met 23,55 % gedaald.

Dat indexcijfer is kleiner dan het ongewogen samengesteld hoeveelheidsindexcijfer omdat de tv's waarvan de hoeveelheidswijziging het grootst is ook de grootste weging hebben.

Merk zoals bij het gewogen samengesteld prijsindexcijfer van Laspeyres op dat als de prijzen van 20x1 sterk verschillen van die van de actuele periode, de prijzen van de basisperiode hier ook weinig representatief zijn voor de actuele periode. Het indexcijfer is dan eventueel minder opportuun.

B Paasche-hoeveelheidsindexcijfer

Het Paasche-hoeveelheidsindexcijfer wordt berekend door de prijzen uit de actuele periode als wegingsfactoren te gebruiken.

De formule wordt in dit geval:

$$I_q^P = \frac{\sum q_t \times p_t}{\sum q_0 \times p_t} \times 100$$

Voorbeeld

We illustreren de formule voor het Paasche-hoeveelheidsindexcijfer op basis van het bovenstaande voorbeeld, maar in plaats van de prijzen van 20x1, gebruiken we nu de prijzen van 20x2:

Tabel 11 Jaarlijkse verkochte hoeveelheden van drie producten en hun prijzen in het actuele jaar 20x2

Product	Hoeveelheid in 20x1	Hoeveelheid in 20x2	Prijs in 20x2
Radio	10 000	8 300	€ 54
Dvd-speler	15 000	14 200	€ 102
Tv	25 000	18 500	€ 550

We berekenen I_q^P:

$$I_q^P = \frac{(8\,300 \times 54) + (14\,200 \times 102) + (18\,500 \times 550)}{(10\,000 \times 54) + (15\,000 \times 102) + (25\,000 \times 550)} \times 100 = \frac{12\,071\,600}{15\,820\,000} \times 100 = 76{,}31$$

De verkochte hoeveelheid in 20x2 daalde nu met 23,69 % ten opzichte van het basisjaar. Het hoeveelheidsindexcijfer van Paasche is kleiner dan dat van Laspeyres omdat de weging van tv's bij het indexcijfer van Paasche groter is dan de weging ervan bij Laspeyres.

C Fisher – hoeveelheidsindexcijfer

De tekortkomingen in verband met de prijswegingscoëfficiënten in de formules van zowel Laspeyres als die van Paasche wordt door het hoeveelheidsindexcijfer van Fisher opgelost. Dat gebeurt opnieuw door een meetkundig gemiddelde te berekenen van beide vorige formules:

$$I_q^F = \sqrt{I_q^L \times I_q^P}$$

Het resultaat voor I_q^F met betrekking tot het jaar 20x2 is dan:

$$I_q^F = \sqrt{76{,}45 \times 76{,}31} = 76{,}38$$

De verkochte hoeveelheid in 20x2 is ten opzichte van 20x1, het basisjaar, gedaald met 23,62 %.

4 Waarde-indexcijfer

Het waarde-indexcijfer meet het samengesteld effect van een hoeveelheidsverandering en een prijsverandering van een goed of een korf van goederen op een bepaald tijdstip ten opzichte van een bepaalde basisperiode.

4.1 Enkelvoudig waarde-indexcijfer

Het enkelvoudig waarde-indexcijfer combineert de prijs- en hoeveelheidsverandering van een goed op een bepaald tijdstip ten opzichte van een basisperiode.

De formule voor dit niet samengesteld indexcijfer is:

$$I_w^e = \frac{q_t \times p_t}{q_0 \times p_0} \times 100$$

Voorbeeld

We passen de formule toe op de radio's die het elektronicabedrijf in 20x1 en 20x2 verkocht heeft:

Tabel 12 Jaarlijkse verkochte hoeveelheden en prijzen van radio's in het basis- en actueel jaar

Product	Hoeveelheid in 20x1	Hoeveelheid in 20x2	Prijs in 20x1	Prijs in 20x2
Radio	10 000	8 300	€ 50	€ 54

Dat geeft voor I_w^e het volgend resultaat in 20x2:

$$I_w^e = \frac{8\,300 \times 54}{10\,000 \times 50} \times 100 = \frac{448\,200}{500\,000} \times 100 = 89,64$$

Het omzetcijfer van de radio's is in 20x2 gedaald met 10,36 % ten opzichte van het basisjaar.

4.2 Samengesteld waarde-indexcijfer

Dit waarde-indexcijfer combineert de hoeveelheids- en prijsverandering van een korf goederen op een bepaald tijdstip ten opzichte van een basisperiode.

De formule voor I_w^s is:

$$I_w^s = \frac{\sum q_t \times p_t}{\sum q_0 \times p_0} \times 100$$

Voorbeeld

In onderstaande tabel hernemen we de hoeveelheden en de prijzen van de artikelen die het elektronicabedrijf verkocht heeft in 20x1 en 20x2:

Tabel 13 Jaarlijkse afzet en prijzen voor de drie producten in 20x1 en 20x2

Product	Hoeveelheid in 20x1	Hoeveelheid in 20x2	Prijs in 20x1	Prijs in 20x2
Radio	10 000	8 300	€ 50	€ 54
Dvd-speler	15 000	14 200	€ 100	€ 102
Tv	25 000	18 500	€ 500	€ 550

We berekenen I^s_w voor 20x2 :

$$I^s_w = \frac{(8\,300 \times 54) + (14\,200 \times 102) + (18\,500 \times 550)}{(10\,000 \times 50) + (15\,000 \times 100) + (25\,000 \times 500)} \times 100 = \frac{12\,071\,600}{14\,500\,000} \times 100 = 83{,}25$$

Het totale omzetcijfer van het elektronicabedrijf is in 20x2 met 16,75 % gedaald ten opzichte van het basisjaar.

5 Wijziging van het basisjaar

In bepaalde gevallen wordt het basisjaar van een reeks indexcijfers gewijzigd. Dat kan bijvoorbeeld gebeuren wanneer de vastgestelde wegingscoëfficiënten van het oude basisjaar op een bepaald ogenblik niet meer representatief zijn. Of als de korf van geconsumeerde goederen ter berekening van een prijsindexcijfer op een bepaald ogenblik niet meer overeenstemt met de korf vastgesteld in een basisjaar, moet er ook overgegaan worden naar een nieuw basisjaar.

6 De schakelverhouding

De schakelverhouding berekent onder meer de prijsverandering van een goed in een bepaalde periode ten opzichte van een vorige periode. Om de schakelverhouding te berekenen, moeten we de prijs van een goed in de actuele periode vergelijken met de prijs ervan in de vorige periode.

De formule voor de schakelverhouding is dan:

$$S_{t,t-1} = \frac{p_t}{p_{t-1}} \times 100$$

Voorbeeld

In de tabel hierna geven we de verkoopprijs weer voor een elektrische wagen verkocht door de onderneming Mesli in de periode 20x1-20x5:

Tabel 14 Verkoopprijs elektrische wagen voor de periode 20x1-20x5

Jaartal	**Verkoopprijs elektrische wagen (euro's)**
20x1	75 000
20x2	80 000
20x3	76 000
20x4	70 000
20x5	67 500

De resultaten van de berekening van de schakelverhoudingen op basis van de gegevens in de vorige tabel, geven we weer in de volgende tabel:

Tabel 15 Schakelverhoudingen 20x2-20x5

Jaartal	**Verkoopprijs elektrische wagen (euro's)**	**Schakelverhouding**
20x1	75 000	
20x2	80 000	106,67
20x3	76 000	95,00
20x4	70 000	92,10
20x5	67 500	96,43

Omdat de prijs voor 20x0 niet gekend is, kunnen we $S_{20x1,20x0}$ niet berekenen.
Uit deze tabel kunnen we onder andere wel afleiden dat de prijs van het voertuig in 20x2 gestegen is met 6,67 % ten opzichte van 20x1 en in 20x5 gedaald is met 3,57 % ten opzichte van 20x4.

7 Economische indexcijfers

Een bekend indexcijfer is de consumptieprijsindex (CPI). Dit indexcijfer meet de evolutie van de levensduurte of van de kostprijs van het levensonderhoud. Deze indicator peilt daarvoor naar de prijsontwikkeling van een korf consumptiegoederen en diensten die representatief is voor het consumptiegedrag van gezinnen.

De gezondheidsindex is op zijn beurt afgeleid uit de CPI. De berekening van dat indexcijfer houdt geen rekening met een aantal goederen zoals alcoholische dranken, tabakswaren en motorbrandstoffen.

De volgende tabel geeft de evolutie van de CPI, de inflatie en de gezondheidsindex in België weer voor de periode van 2000 tot 2011:

Tabel 16 Consumptieprijsindexen en inflatie in België

2004 = 100	CPI	Inflatie (%)	Gezondheidsindex
2000	92,6	2,5	92,8
2001	94,9	2,5	95,2
2002	96,4	1,6	97,0
2003	97,9	1,6	98,4
2004	100,0	2,1	100,0
2005	102,8	2,8	102,10
2006	104,6	1,8	104,0
2007	106,5	1,8	105,8
2008	111,3	4,5	110,3
2009	111,3	0,0	110,9
2010	113,7	2,2	112,8
2011	117,7	3,5	116,1

Voor het jaar 2011 zien we dat de CPI steeg tot 117,7 en dat de inflatie in dat jaar overeenstemde met een groeipercentage van 3,5 %. De inflatie op basis van de gezondheidsindex is uiteraard verschillend omdat de evolutie van de levensduurte wordt berekend rekening houdend met een andere korf van consumptiegoederen en diensten.

Een tweede belangrijk indexcijfer in de economie is het bruto binnenlands product per capita-indexcijfer (bbpci). Dat indexcijfer geeft de evolutie van de welvaart van de bevolking in een land weer.

Onderstaande tabel toont de gegevens voor België gedurende de periode van 2005 tot 2014 van het bbp per inwoner en het bbpci:

Tabel 17 Bbp per inwoner en bbpci in België

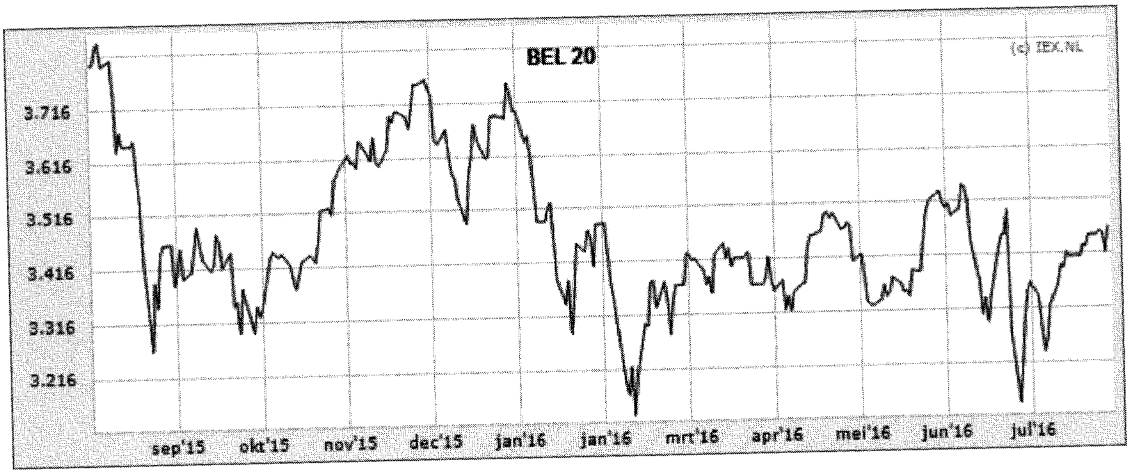
2005 = 100	bbp per inwoner (in euro koopkrachtpariteiten)	Bbpci(1)
2005	27 632	100,00
2006	28 621	103,58
2007	29 723	107,57
2008	29 658	107,33
2009	28 405	102,80
2010	30 305	109,67
2011	31 187	112,87
2012	31 705	114,74
2013	31 615	114,41
2014	32 301	116,90

Bron: Eurostat, HERMREG
(1) Eigen berekening

Voor het jaar 2014 stellen we vast dat de gemiddelde welvaart in België met 16,90 % is gestegen ten opzichte van het basisjaar 2005.

Ten slotte hebben we het ook nog even over het voornaamste beursindexcijfer in de Belgische economie: de BEL-20. Deze index is de pols van de Belgische economie en bestaat uit de twintig belangrijkste aandelen.

De eerste berekening van deze index gebeurde op 30 maart 1990 en de waarde van de korf werd toen gelijkgesteld aan 1000 punten. In onderstaande grafiek van de Bel-20 kunnen we aflezen dat de index op 1 augustus 2016 op 3 448,85 punten staat. De korf van twintig toonaangevende aandelen steeg 244,89 % in waarde in een periode van meer dan 26 jaren.



Bron: Beursduivel.be

hoofdstuk 8

Nog meer statistische functies

1 AANTAL (COUNT) en AANTALARG (COUNTA)

De functie AANTAL telt het aantal cellen dat getallen bevat.

In Figuur 31 hebben we drie kolommen met een aantal getallen gevuld en cel B18 geselecteerd om het resultaat van de AANTAL-functie te tonen.
We klikken op de f_x-knop en openen het venster 'Functie invoegen'. We selecteren de functie 'AANTAL' en openen het venster 'Functieargumenten'. In het vak 'Waarde 1' voeren we de matrix B1:B16 in, in het vak 'Waarde 2' het bereik C1:C6 en in het vak 'Waarde 3' het bereik D1:D7. We klikken op 'OK' en in ons werkblad vinden we in cel B18 het aantal cellen met getallen, namelijk 29.

B18 =AANTAL(B1:B16;C1:C6;D1:D7)

	A	B	C	D	E	F
1		58	80	50		
2		63	70	65		
3		71	65	90		
4		77	73	43		
5		60	21	22		
6		89	18	19		
7		56		35		
8		81				
9		73				
10		62				
11		95				
12		48				
13		54				
14		69				
15		80				
16		67				
17						
18	AANTAL	29				
19						

Figuur 31 Berekening van het aantal cellen met getallen

De functie AANTALARG telt zowel cellen die tekst bevatten als cellen die getallen bevatten. In Figuur 32 illustreren we dat.

B18 =AANTALARG(B1:B16;C1:C8;D1:D7)

	A	B	C	D	E	F	G
1		58	80	50			
2		63	70	65			
3		71	65	90			
4		77	73	43			
5		60	21	22			
6		89	18	19			
7		56	ja	35			
8		81	nee				
9		73					
10		62					
11		95					
12		48					
13		54					
14		69					
15		80					
16		67					
17							
18	AANTALARG	31					
19							

Figuur 32 Berekening van het aantal cellen dat getallen of tekst bevat

2 AANTAL.LEGE.CELLEN (COUNTBLANK)

Met deze functie wordt het aantal lege cellen geteld in een bepaald bereik.

We keren terug naar de gegevens van Figuur 32 en willen weten hoeveel lege cellen het bereik B1:D16 bevat. Je gaat als volgt te werk:

> klik op de f_x-knop en selecteer in het venster 'Functie invoegen' de functie 'AANTAL.LEGE.CELLEN';

> nadien open je het venster 'Functieargumenten', je voert in het vak 'Bereik' de matrix B1:D16 in, klikt op 'OK' en we lezen in de geselecteerde cel B18 in het werkblad het aantal lege cellen, zijnde 17.

Het resultaat zien we in Figuur 33.

B18 =AANTAL.LEGE.CELLEN(B1:D16)

	A	B	C	D	E
1		58	80	50	
2		63	70	65	
3		71	65	90	
4		77	73	43	
5		60	21	22	
6		89	18	19	
7		56	ja	35	
8		81	nee		
9		73			
10		62			
11		95			
12		48			
13		54			
14		69			
15		80			
16		67			
17					
18	AANTAL.LEGE.CELLEN	17			
19					

Figuur 33 Berekening van het aantal lege cellen in een bereik

3 AANTAL.ALS (COUNTIF)

Dit is een functie die in een bereik de cellen telt die voldoen aan een bepaald criterium.

In Figuur 34 berekenen we in cel B18 het aantal cellen dat voldoet aan het criterium ≤ 60. Daartoe werken we de volgende stappen af:

> we klikken op de f_x-knop en openen het venster 'Functie invoegen';
> we selecteren 'AANTAL.ALS', klikken op 'OK' en het venster 'Functieargumenten' verschijnt;
> we voeren het bereik B1:D16 in, en in het vak 'Criterium' zetten we ≤ 60;
> we klikken op 'OK' en in de cel B18 verschijnt het aantal, dat is 12.

B18 | =AANTAL.ALS(B1:D16;"<=60")

	A	B	C	D	E	F
1		58	80	50		
2		63	70	65		
3		71	65	90		
4		77	73	43		
5		60	21	22		
6		89	18	19		
7		56		35		
8		81				
9		73				
10		62				
11		95				
12		48				
13		54				
14		69				
15		80				
16		67				
17						
18	AANTAL.ALS	12				
19						

Figuur 34 Berekening van het aantal cellen dat voldoet aan het criterium ≤ 60

4 MAX (MAX) en MIN (MIN)

Excel heeft twee statistische functies die als resultaat hetzij het kleinste getal hetzij het grootste getal geven in een bereik met waarden.

We geven de stappen voor de MAX-functie en gebruiken daarvoor de bereiken B1:B16, C1:C6 en D1:D7 in Figuur 35:

> klik op de f_x-knop en het venster 'Functie invoegen' verschijnt;
> selecteer de functie 'MAX' die we uitvoeren in de geselecteerde cel B17. Het venster 'Functieargumenten' wordt geopend en in de vakken 'Getal 1', 'Getal 2' en 'Getal 3' voer je respectievelijk de bereiken B1:B16, C1:C6 en D1:D7 in;
> sluit het venster met 'OK' en in cel B17 verschijnt het grootste getal: 95.

De MIN-functie werkt volgens dezelfde stappen als de MAX-functie.

B17 =MAX(B1:B16;C1:C6;D1:D7)

	A	B	C	D	E	F
1		58	80	50		
2		63	70	65		
3		71	65	90		
4		77	73	43		
5		60	21	22		
6		89	18	19		
7		56		35		
8		81				
9		73				
10		62				
11		95				
12		48				
13		54				
14		69				
15		80				
16		67				
17	MAX	95				
18						

Figuur 35 Berekening van de grootste waarde in een bereik

5 SCHEEFHEID (SKEW)

Een histogram of kolommendiagram is symmetrisch als de twee helften van de grafiek elkaars spiegelbeeld zijn. In dat geval is de scheefheid $\bar{x} - Mo = 0$.[2]

Als een histogram asymmetrisch is, of m.a.w. de waarnemingen niet gelijk verdeeld zijn, onderscheiden we twee mogelijkheden:

1 *scheefheid naar rechts* betekent dat de grafiek naar rechts een lange staart heeft. De scheefheid is dan positief of $\bar{x} - Mo > 0$;
2 *scheefheid naar links* wijst op een grafiek met een lange staart naar links waarbij $\bar{x} - Mo < 0$.

Een maatstaf voor de scheefheid is de scheefheidscoëfficiënt van Pearson. Deze is:

- voor een steekproef: $S_p = \dfrac{\bar{x} - Mo}{s}$
- voor een populatie: $S_p = \dfrac{\mu - Mo}{\sigma}$

2 In geval van een populatie is de formule $\mu - Mo = 0$.

We illustreren in Figuur 36 het gebruik van de SCHEEFHEID-functie en dit voor een verdeling op basis van een steekproef:

> we geven in het bereik A1:D10 een aantal getallen in en selecteren D12 voor de berekening van de SCHEEFHEID;
> we klikken op de f_x-knop en het venster 'Functie invoegen' wordt geopend, waarna we de functie 'SCHEEFHEID' selecteren;
> in het venster 'Functieargumenten' vullen we in het vak 'Getal 1' ons gegevensbereik A1:D10 in;
> we klikken op 'OK' en we lezen in cel D12 dat de scheefheid positief is of +0,563449.

D12 | f_x =SCHEEFHEID(A1:D10)

	A	B	C	D	E	F
1	15	28	63	40		
2	16	22	69	24		
3	33	17	35	13		
4	32	19	38	17		
5	41	53	23	25		
6	66	54	26	38		
7	11	56	18	39		
8	35	40	17	22		
9	60	45	7	17		
10	30	42	5	9		
11						
12			SCHEEFHEID	0,563		
13						

Figuur 36 Berekening van de scheefheid

hoofdstuk 9

Frequentieverdelingen

Er zijn twee manieren om met Excel een frequentieverdeling op te maken. De eerste manier is met behulp van een werkbladfunctie, de tweede maakt gebruik van een gegevensanalysefunctie.

1 INTERVAL (FREQUENCY)

Met deze werkbladfunctie maken we een frequentieverdeling. Dat betekent dat we intervallen of klassen creëren waarin we onze gegevens zullen indelen.

In het voorbeeld in Figuur 37 gebruiken we opnieuw de gegevens waarmee we de scheefheid van een grafiek hebben berekend. We werken als volgt:

> in een kolom 'Interval' voeren we de verschillende intervallen in;
> in cel G1 voeren we de titel 'Frequentie' in en we selecteren G2:G9 waarin we de frequenties willen presenteren;
> we klikken op de f_x-knop, openen het venster 'Functie invoegen', selecteren de functie 'INTERVAL' en klikken op 'OK';
> in het venster 'Functieargumenten' voeren we in het vak 'Gegevensmatrix' het bereik A1:D10 in en in het vak 'Interval_verwijzing' het bereik F2:F8;
> we drukken terzelfdertijd op de toetsen CTRL, SHIFT en OK.

In het werkblad zien we in het bereik G2:G9 de gewenste frequenties.

G2 | {=INTERVAL(A1:D10;F2:F8)}

	A	B	C	D	E	F	G	H
1	15	28	63	40		Interval	Frequentie	
2	16	22	69	24		10	3	
3	33	17	35	13		20	10	
4	32	19	38	17		30	8	
5	41	53	23	25		40	9	
6	66	54	26	38		50	3	
7	11	56	18	39		60	4	
8	35	40	17	22		70	3	
9	60	45	7	17			0	
10	30	42	5	9				
11								

Figuur 37 Opmaken van een frequentieverdeling

2 Histogram (HISTOGRAM)

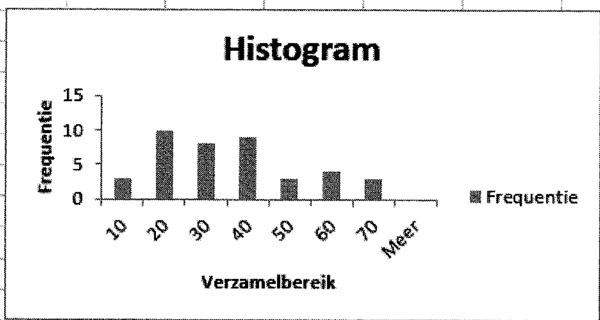
Met de gegevensanalysefunctie Histogram bereiken we hetzelfde resultaat.

Met dezelfde gegevens in het bereik A1:D10 en met dezelfde intervalindeling van Figuur 37 gaan we als volgt te werk:

> we klikken in het lint op het tabblad 'Gegevens' en selecteren de keuzelijst 'Gegevensanalyse'. In het dialoogvenster dat verschijnt kiezen we de functie 'Histogram'. Het dialoogvenster 'Histogram' wordt geopend;
> in het vak Invoerbereik voeren we het gegevensbereik A1:D10 in en in het Verzamelbereikvak zetten we het intervalbereik F2:F8;
> we selecteren de uitvoeroptie 'Uitvoerbereik' en voeren cel A12 in door in het werkblad te klikken op cel A12. Deze cel verwijst naar de linkerbovenhoek van de uitvoertabel;
> verder selecteren we de uitvoeroptie 'Grafiek maken' en klikken op 'OK'.

In Figuur 38 zien we het resultaat. Door te klikken in het Grafiekgebied en door verder te klikken op het hoekpunt in de rechterbenedenhoek en de muisknop ingedrukt te houden, kunnen we het grafiekkader vergroten als we de muisaanwijzer naar de benedenrand van het scherm verplaatsen.
Het histogram is – zoals we zien – consistent met de eerder berekende positieve scheefheid.

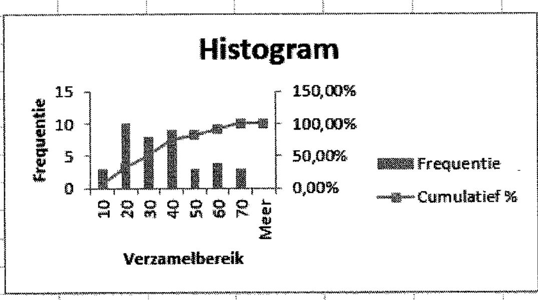
	A	B	C	D	E	F	G	H	I	J
1	15	28	63	40		Interval				
2	16	22	69	24		10				
3	33	17	35	13		20				
4	32	19	38	17		30				
5	41	53	23	25		40				
6	66	54	26	38		50				
7	11	56	18	39		60				
8	35	40	17	22		70				
9	60	45	7	17						
10	30	42	5	9						
11										
12	*Verzamelbereik*	*Frequentie*								
13	10	3								
14	20	10								
15	30	8								
16	40	9								
17	50	3								
18	60	4								
19	70	3								
20	Meer	0								
21										
22										
23										



Figuur 38 Het histogram en de frequentieverdeling in klassen

Er bestaat nog een andere uitvoeroptie in het Histogramvenster, met name Cumulatief percentage. Deze optie levert een bijkomende kolom op met de cumulatieve relatieve frequenties en een cumulatieve relatieve frequentielijn. Dat wordt getoond in Figuur 39 hieronder.

	A	B	C	D	E	F	G	H	I	J	K
1	15	28	63	40		Interval					
2	16	22	69	24		10					
3	33	17	35	13		20					
4	32	19	38	17		30					
5	41	53	23	25		40					
6	66	54	26	38		50					
7	11	56	18	39		60					
8	35	40	17	22		70					
9	60	45	7	17							
10	30	42	5	9							
11											
12	*Verzamelbereik*	*Frequentie*	*Cumulatief %*								
13	10	3	7,50%								
14	20	10	32,50%								
15	30	8	52,50%								
16	40	9	75,00%								
17	50	3	82,50%								
18	60	4	92,50%								
19	70	3	100,00%								
20	Meer	0	100,00%								
21											
22											
23											



Figuur 39 Histogram met toepassing van de uitvoeroptie Cumulatief percentage

3 Beschrijvende statistiek-functie

Figuur 40 geeft een overzicht van de behaalde trimestriële resultaten door een groep studenten op een hogeschool.

	A	B	C	D
1	Trim I	Trim II	Trim III	
2	60	56	63	
3	80	66	69	
4	66	73	70	
5	74	89	76	
6	82	63	46	
7	66	54	63	
8	55	56	59	
9	70	80	41	
10	90	86	45	
11	65	77	49	
12				

Figuur 40 Behaalde trimestriële resultaten van studenten op een hogeschool

Door gebruik te maken van de functie Beschrijvende statistiek krijgen we voor de drie trimesters telkens een verzameling statistische grootheden die de trimestriële resultaten beschrijven.

Om de functie Beschrijvende statistiek te gebruiken, ga je als volgt te werk:

> klik in het lint op het tabblad 'Gegevens' en kies vervolgens de keuzelijst 'Gegevensanalyse' om het venster 'Gegevensanalyse' te openen;
> je kiest de functie 'Beschrijvende statistiek', klikt op 'OK' en opent het dialoogvenster 'Beschrijvende statistiek';
> in het vak Invoerbereik voer je de gegevensmatrix A1:C11 in;
> je klikt op 'Groeperen per kolom' om aan te geven dat de gegevens in kolommen zijn gepresenteerd. Je selecteert 'Labels in de eerste rij' omdat de gegevensmatrix kolomtitels bevat;
> bij de Uitvoeropties selecteer je 'Nieuw tabblad werkblad' en 'Samenvattingsinfo';
> we klikken op 'OK' en in Figuur 41 zien we het nieuwe tabblad dat een overzicht geeft van waarden van een reeks statistische grootheden per trimester.

	A	B	C	D	E	F	G
1	*Trim I*		*Trim II*		*Trim III*		
2							
3	Gemiddelde	70,8	Gemiddelde	70	Gemiddelde	58,1	
4	Standaardfout	3,392148	Standaardfout	4,06612	Standaardfout	3,833913	
5	Mediaan	68	Mediaan	69,5	Mediaan	61	
6	Modus	66	Modus	56	Modus	63	
7	Standaarddeviatie	10,72691	Standaarddeviatie	12,8582	Standaarddeviatie	12,1239	
8	Steekproefvariantie	115,0667	Steekproefvariantie	165,3333	Steekproefvariantie	146,9889	
9	Kurtosis	-0,43227	Kurtosis	-1,51789	Kurtosis	-1,49837	
10	Scheefheid	0,413348	Scheefheid	0,152485	Scheefheid	-0,06405	
11	Bereik	35	Bereik	35	Bereik	35	
12	Minimum	55	Minimum	54	Minimum	41	
13	Maximum	90	Maximum	89	Maximum	76	
14	Som	708	Som	700	Som	581	
15	Aantal	10	Aantal	10	Aantal	10	
16							

Figuur 41 Uitvoer van de functie Beschrijvende statistiek

hoofdstuk 10

Kansrekening

1 Experiment

Een experiment kunnen we definiëren als een toevallig verschijnsel waarvan de uitkomsten onzeker zijn of onvoorzienbaar zijn. Een voorbeeld is het opwerpen van een dobbelsteen, waarbij we niet kunnen voorzien welk zijvlak we zullen uitkomen.

Een uitkomst van een experiment wordt in de statistiek aangeduid met de term 'enkelvoudige gebeurtenis'. Het opwerpen van zes ogen met een dobbelsteen is bijvoorbeeld een enkelvoudige gebeurtenis omdat er maar op één manier zes ogen opgeworpen kunnen worden, en dat is door het zijvlak met de zes ogen te gooien.

Alle uitkomsten van een experiment vormen de uitkomstenverzameling van dat experiment. De uitkomsten 1, 2, 3, 4, 5 en 6 vormen de uitkomstenverzameling van het opwerpen van een dobbelsteen.

De kans van een enkelvoudige gebeurtenis van een experiment is, als de kans van iedere gebeurtenis gelijk is:

$$P\text{ (enkelvoudige gebeurtenis)} = \frac{1}{\text{aantal uitkomsten in de uitkomstenverzameling}}$$

De kans op het gooien van zes ogen is dus:

$$P\ (6) = \frac{1}{6}$$

Een samengestelde gebeurtenis van een experiment, bijvoorbeeld het opwerpen van een even aantal ogen, is een gebeurtenis waaraan meer dan één uitkomst van een experiment aan voldoet. De kans van zo'n gebeurtenis is:

$$P(\text{samengestelde gebeurtenis}) = \Sigma P(u)$$

De kans op een samengestelde gebeurtenis is dus gelijk aan de som van de kansen van alle uitkomsten (= u) die voldoen aan die samengestelde gebeurtenis.

De kans op het werpen van een even aantal ogen met een dobbelsteen is dus:

$$P(\text{even aantal ogen}) = \frac{3}{6}$$

Ten slotte is de som van de kansen van alle uitkomsten van een experiment altijd gelijk aan 1.

2 Combinaties van gebeurtenissen

2.1 De unie

De unie van gebeurtenis A en gebeurtenis B is de verzameling die alle uitkomsten bevat die deel uitmaken van A of B of van allebei. Wiskundig wordt de unie voorgesteld door $A \cup B$, waarbij het symbool $\cup$ voor de unie staat.

Nemen we als voorbeeld het experiment van het opwerpen van drie muntstukken. We bepalen A als het opwerpen van twee keer kruis en B als het opwerpen van drie keer munt. Wat is dan de uitkomstenverzameling van $A \cup B$?
Eerst bepalen we de uitkomstenverzameling van ons experiment, nl. het opwerpen van drie muntstukken. Na het opstellen van een kansboom voor ons toevalsverschijnsel, vinden we als uitkomstenverzameling:

V = (KKK, KKM, KMK, MKK, KMM, MKM, MMK, MMM)

of

$A \cup B$ = (KKM, KMK, MKK, MMM)

Aangezien A en B geen gemeenschappelijke uitkomsten hebben, is:

$$P(A \cup B) = P(A) + P(B)$$

of

$$P(A \cup B) = \frac{3}{8} + \frac{1}{8} = \frac{4}{8} = \frac{1}{2}$$

Als A en B wel gemeenschappelijke uitkomsten hebben, dan mogen we voor de kansberekening de gemeenschappelijke uitkomsten geen twee keer in aanmerking nemen. Dat vermijden we door de bovenstaande formule als volgt aan te passen:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$\cap$ staat voor doorsnede.

Bepalen we A als het opwerpen van één keer kruis en B als het opwerpen van ten hoogste tweemaal kruis, dan zijn hun uitkomstenverzamelingen:

A = {KMM, MKM, MMK}
B = {KMM, MKM, MMK, KKM, KMK, MKK, MMM}

De kans van A unie B is dan:

$$P(A \cup B) = \frac{3}{8} + \frac{7}{8} - \frac{3}{8} = \frac{7}{8}$$

2.2 De doorsnede

De doorsnede van de gebeurtenis A en de gebeurtenis B is de verzameling van de uitkomsten die zowel tot A als tot B behoren. Wiskundig wordt de doorsnede door $A \cap B$ voorgesteld.

Als we de gebeurtenis A nu omschrijven als het gooien van minstens twee keer munt en B als het werpen van twee keer munt, dan zijn de uitkomstenverzamelingen:

A = {MMK, MKM, KMM, MMM}
B = {MMK, MKM, KMM}

Omdat A en B afhankelijke gebeurtenissen zijn – de kans van B is afhankelijk van het zich voordoen van A en omgekeerd – is de kansformule voor $A \cap B$:

$$P(A \cap B) = P(A) \times P(B/A) = P(A) \times \frac{P(A \cap B)}{P(A)}$$

of $P(A \cap B) = \frac{4}{8} \times \frac{3/8}{4/8} = \frac{3}{8}$

Het schuine streepje in P(B/A) lezen we als 'gegeven zijnde dat'. In dit voorbeeld wil dat zeggen dat A zich heeft voorgedaan.

Gebeurtenissen kunnen ook onafhankelijk zijn. Nemen we bijvoorbeeld het experiment van het opgooien van twee dobbelstenen. De uitkomstenverzameling V is:

V = {(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6), (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)}

Als we de gebeurtenis A definiëren als het opwerpen van een even aantal ogen bij de eerste dobbelsteen en B als het opwerpen van een oneven aantal ogen bij de tweede dobbelsteen, dan zijn de uitkomstenverzamelingen van A en B:

$A = \{2, 4, 6\}$
$B = \{1, 3, 5\}$

Aangezien A en B onafhankelijk zijn – de uitkomst van B is niet afhankelijk van de uitkomst van A en omgekeerd – is de kansformule voor $A \cap B$:

$$P(A \cap B) = P(A) \times P(B)$$

of

$$P(A \cap B) = \frac{3}{6} \times \frac{3}{6} = \frac{9}{36} = \frac{1}{4}$$

3 Permutaties

Stel dat een sportploeg vijf leden heeft en men ons vraagt op hoeveel verschillende manieren we die leden in een groep van vijf kunnen rangschikken, wetend dat als men de volgorde verandert we een nieuwe groep hebben. Het antwoord op die vraag is dat we voor de keuze van het eerste lid in de groep 5 mogelijkheden hebben, voor het tweede lid 4 mogelijkheden, voor het derde lid 3 mogelijkheden, voor het vierde lid 2 mogelijkheden en ten slotte voor het vijfde lid 1 mogelijkheid. Het aantal groepen dat we kunnen vormen, is:

$$5 \times 4 \times 3 \times 2 \times 1 = 120$$

De formule voor dit soort berekening is:

$$P(n,n) = \frac{n!}{(n-n)!} = \frac{n!}{0!} = n!$$

waarbij $n! = n \times (n-1) \times (n-2) \ldots \times 1$*

Deze formule levert ons het aantal groepen of permutaties op van n elementen uit een groep van n elementen.
n! lezen we als n-faculteit en per definitie is $0! = 1$.

* n! = het product van alle natuurlijke getallen van n tot en met 1.

We kunnen nog een stap verder gaan. Als we uit onze ploeg een kapitein en een onderkapitein moeten kiezen, dan rijst de vraag hoeveel permutaties we kunnen vormen. Voor de keuze van de kapitein zijn er 5 mogelijkheden en voor de keuze van de onderkapitein zijn er 4 mogelijkheden, m.a.w. $5 \times 4 = 20$ mogelijkheden in totaal. Algemeen geformuleerd schrijven we:

$$P(n, r) = \frac{n!}{(n - r)!}$$

Deze formule geeft ons het totale aantal groepen of permutaties van r elementen uit een groep van n elementen.

In ons voorbeeld krijgen we dan:

$$P(5,2) = \frac{5!}{(5 - 2)!} = \frac{5!}{3!} = 20$$

4 Combinaties

Wanneer groepen alleen kunnen verschillen door de opgenomen elementen en niet door hun volgorde te wijzigen, dan duiden we elk van die groepen aan met de term 'combinatie'.

De algemene formule voor de berekening van het aantal combinaties is:

$$\binom{n}{r} = \frac{n!}{r!(n - r)!}$$

Deze formule geeft ons het aantal combinaties van r elementen uit een groep van n elementen, waarbij twee groepen als verschillend worden beschouwd als ze minstens in één van de opgenomen elementen verschillen.

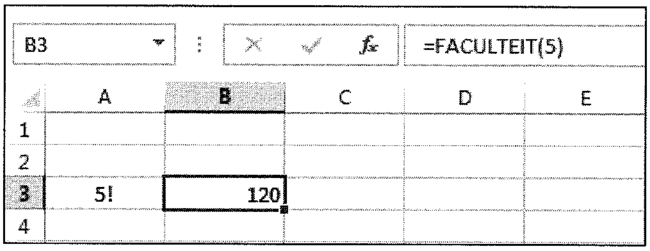
Veronderstel dat we uit onze sportploeg van 5 leden een kapitein en een onderkapitein moeten kiezen en we willen weten hoeveel combinaties van 2 sportleden er mogelijk zijn. Als we de formule toepassen, dan vinden we:

$$\binom{5}{2} = \frac{5!}{2!\,(5 - 2)!} = \frac{5 \times 4}{2} = 10$$

5 FACULTEIT – PERMUTATIES – COMBINATIES (FACT – PERMUT – COMBIN)

De eerste Excelfunctie FACULTEIT (FACT) berekent de faculteit van een getal. Ondertussen weten we dat bijvoorbeeld de faculteit van het getal n gelijk is aan n × (n – 1) × (n – 2) … × 1. De functie gebruiken we als volgt:

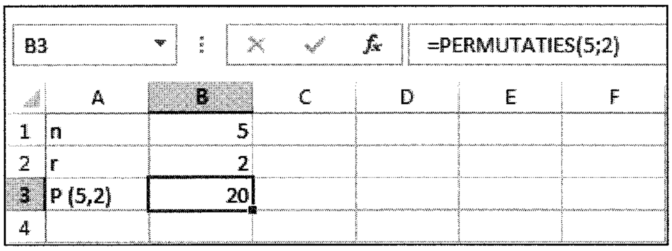
> we selecteren in het werkblad een cel, bijvoorbeeld B3, waarin we het resultaat van de faculteitberekening willen tonen;
> we klikken op de f_x-knop en openen het venster 'Functie invoegen', selecteren de categorie 'Wiskunde en trigonometrie' en daarin kiezen we de functie 'FACULTEIT';
> in het geopende venster 'Functieargumenten' voeren we in het vak 'Getal' het cijfer in waarvan we de faculteit willen berekenen, bijvoorbeeld het getal 5;
> we sluiten het venster 'Functieargumenten' en het antwoord '120' verschijnt in cel B3. We zien dat in Figuur 42.



Figuur 42 De FACULTEIT-functie

De functie PERMUTATIES (PERMUT) illustreren we met het voorbeeld van de sportploeg, waarbij we ons afvroegen hoeveel permutaties van 2 elementen uit een groep van 5 elementen we konden vormen. We moeten de volgende stappen uitvoeren:

> in het werkblad van Figuur 43 selecteren we cel B3 voor het antwoord op ons probleem;
> We klikken op de f_x-knop en kiezen in het dialoogvenster 'Functie invoegen' de statistische functie 'PERMUTATIES';
> in het geopende venster 'Functieargumenten' vullen we in het vak 'Getal' het totale aantal elementen n in. In ons voorbeeld is dat 5;
> in het vak 'Aantal-gekozen' vullen we het aantal elementen r in elke permutatie in. In ons voorbeeld is dat 2;
> we sluiten het venster 'Functieargumenten' en het antwoord verschijnt in cel B3, zijnde 20.



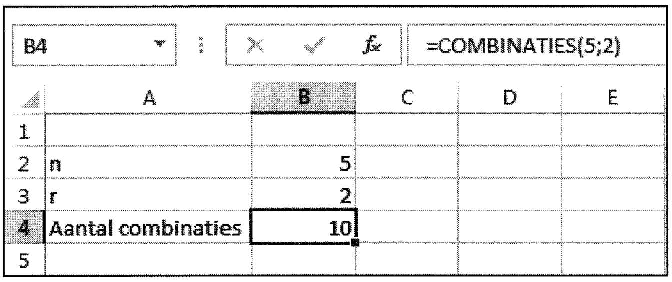
B3 =PERMUTATIES(5;2)

	A	B	C	D	E	F
1	n	5				
2	r	2				
3	P (5,2)	20				
4						

Figuur 43 De functie PERMUTATIES

Ten slotte kunnen we met de functie COMBINATIES (COMBIN) het aantal verschillende groepen van 2 sportleden – de kapitein en de onderkapitein – uit een ploeg van 5 leden berekenen:

> in Figuur 44 selecteren we in het werkblad cel B4 voor het resultaat van de functie;
> we klikken op de f_x-knop en openen het venster 'Functie invoegen';
> we selecteren de categorie 'Wiskunde en trigonometrie', kiezen dan de functie 'COMBINATIES' en klikken op 'OK';
> het venster 'Functieargumenten' verschijnt en in het vak 'Getal' voeren we n in, dat gelijk is aan 5;
> in het vak 'Aantal – gekozen' voeren we r in of het aantal elementen in elke combinatie, namelijk 2;
> we sluiten het venster 'Functieargumenten' en het resultaat, 10 combinaties, verschijnt in cel B4.



B4 =COMBINATIES(5;2)

	A	B	C	D	E
1					
2	n	5			
3	r	2			
4	Aantal combinaties	10			
5					

Figuur 44 De functie COMBINATIES

hoofdstuk 11

Discrete kansverdelingen

1 De binomiale kansverdeling

1.1 Gebruik

Een experiment waarbij de mogelijke uitkomsten van een toevalsvariabele worden ingedeeld als een succes of een mislukking, is een binomiaal experiment. Een voorbeeld van een binomiaal experiment is een muntstuk meerdere keren opwerpen. Het opwerpen van kruis kan dan bijvoorbeeld beschouwd worden als een succes en het opwerpen van munt als een mislukking. Daarbij kan men zich dan de vraag stellen wat de kans is om bij n worpen 0, 1, 2, ..., n successen te hebben. Het aantal successen noemt men in ons voorbeeld de binomiale toevalsvariabele x die discreet van aard is.

Om de kansen van de mogelijke uitkomsten van x te bepalen, moeten we voor ogen houden dat de uitkomst van een worp van een muntstuk onafhankelijk is van de uitkomsten van de vorige worpen en dat bij iedere worp de kans op succes en de kans op mislukking constant zijn.

De formule waarmee we de kansverdeling van een binomiale toevalsvariabele x kunnen berekenen, ziet er als volgt uit:

$$P(x) = \binom{n}{x} \times p^x \times (1 - p)^{n-x}$$

waarbij: n = aantal uitvoeringen van het experiment
x = aantal successen
p = kans op succes
1 − p = kans op mislukking
n − x = aantal mislukkingen
$\binom{n}{x}$ = aantal manieren van x successen bij n uitvoeringen van het experiment

De berekening van de kans op het werpen van viermaal kruis bij 6 worpen wordt dan:

$$P(x = 4) = \binom{6}{4} \times \left(\frac{1}{2}\right)^4 \times \left(\frac{1}{2}\right)^2$$

$$= 0{,}234$$

1.2 BINOM. VERD (BINOMDIST)

Aan de hand van het voorbeeld dat we hierboven beschreven, tonen we nu het gebruik aan van de functie BINOM. VERD:

> selecteer in het werkblad van Figuur 45 de cel B4;
> klik op de f_x-knop, open het venster 'Functie invoegen', kies de functie 'BINOM. VERD' en klik op 'OK';
> in het venster 'Functieargumenten' voer je in het vak 'Aantal – gunstig' het getal 4 in;
> in het vak 'Experimenten' vul je het aantal worpen in, zijnde 6;
> in het vak 'Kans – gunstig' voer je de kans op succes in, in dit voorbeeld is dat 0,5;
> in het vak 'Cumulatief' vul je 'ONWAAR' in als we de exacte kans willen kennen van het aantal successen ingevuld in het vak 'Aantal – gunstig'. Vul je 'WAAR' in, dan krijg je als resultaat de kans op 4 of minder successen. Wij vullen 'ONWAAR' in;
> het antwoord 0,234375 verschijnt in cel B4 na het sluiten van het venster.

B4 | =BINOM.VERD(4;6;0,5;ONWAAR)

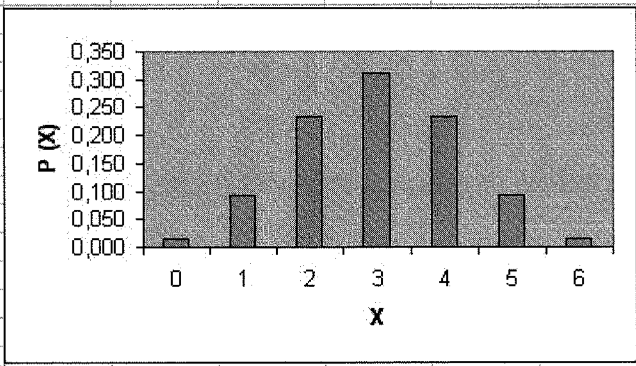
	A	B	C	D	E	F
1	n	6				
2	X	4				
3	p	0,5				
4	P (X = 4)	0,234375				
5						

Figuur 45 De BINOM.VERD-functie

In Figuur 46 berekenen we de binomiale verdeling van ons experiment met de functie BINOM.VERD. Het bijbehorende kolommendiagram tekenen we als volgt:

> selecteer de gegevens B5:B11, klik op 'Invoegen' en selecteer in de keuzelijst 'Kolom- of staafdiagram invoegen' het subtype 'Gegroepeerde kolom';
> selecteer dan in de groep 'Gegevens', 'Gegevens selecteren'. Het dialoogvenster 'Gegevensbron selecteren' verschijnt. Klik op 'Bewerken' van 'Horizontale aslabels (categorieën)' en geef in het vak 'Aslabelbereik' het bereik A5:A11 op, zodat de juiste waarden voor de X-variabele langs de horizontale as verschijnen;
> klik vervolgens op het tabblad 'Ontwerpen', dan op de keuzelijst 'Grafiekonderdelen toevoegen', selecteer 'Astitels' en kies de optie 'Primair horizontaal'. Vul in het astitelvak dat verschijnt de letter X in. Selecteer opnieuw 'Astitels' om de verticale as van een titel te voorzien, kies 'Primair verticaal' en vul P(X) in het titelvak in;
> verwijder ten slotte nog de grafiektitel en de rasterlijnen.

	A	B
1	p	0,5
2	n	6
3		
4	X	P (X)
5	0	0,016
6	1	0,094
7	2	0,234
8	3	0,313
9	4	0,234
10	5	0,094
11	6	0,016



Figuur 46 Binomiale kansverdeling van X aantal keren kruis bij het zes keer opwerpen van een muntstuk

We hebben er bij de bespreking van de functie BINOM. VERD al op gewezen dat we in het vak 'Cumulatief' ook de waarde 'WAAR' kunnen invoeren. Als we dat in ons voorbeeld doen, vinden we in Figuur 47 als uitkomst een kans van 0,891 op ten hoogste 4 successen.

B4 — =BINOM.VERD(4;6;0,5;WAAR)

	A	B	C	D	E	F	G
1	p	0,5					
2	n	6					
3	X	<=4					
4	P (x<=4)	0,890625					
5							

Figuur 47 Berekening van de kans op ten hoogste 4 keren kruis bij 6 worpen

1.3 Testen van een hypothese met de binomiale kansverdeling

Om een hypothese te testen met de binomiale kansverdeling, formuleren we eerst onze mening over de kans op succes in een nulhypothese. Dan voeren we het binomiale experiment n keren uit en noteren het aantal successen. We berekenen de kans dat we zoveel successen of meer kunnen behalen als H_0 waar is. Als deze kans klein is, verwerpen we H_0.

Wat we doen, is een conclusie trekken omtrent een populatieparameter. In ons voorbeeld is deze parameter de kans op een succes in de populatie van N uitvoeringen van ons experiment. We nemen als voorbeeld een muntstuk waarvan we de zuiverheid willen testen. We veronderstellen dat het niet zuiver is, omdat volgens ons het muntstuk de neiging vertoont meer kruis op te leveren dan normaal.

We beschouwen 12 worpen van het muntstuk en stellen vast dat we 9 keer kruis werpen. De hypothesen zijn:

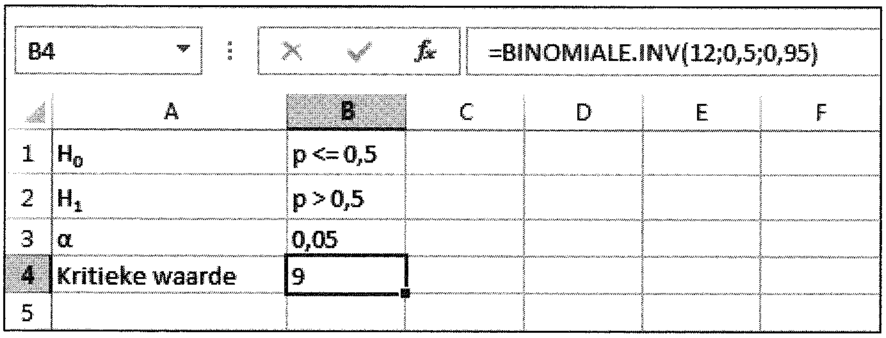
H_0: p ≤ 0,5
H_1: p > 0,5
α = 0,05

p = kans op succes in een populatie van N uitvoeringen van een experiment.

Om deze hypothesen te testen, gebruiken we de statistische Excelfunctie BINOMIALE.INV (BINOM.INV) Deze functie levert een kritieke waarde op van het aantal successen. Als het aantal waargenomen successen gelijk is aan of groter is dan de kritieke waarde, verwerpen we H_0.
We doorlopen de stappen van deze functie voor ons voorbeeld:

- we selecteren in het werkblad van Figuur 48 cel B4 om het antwoord te laten verschijnen;
- we klikken op de f_x-knop en openen het venster 'Functie invoegen';
- in het venster 'Functie invoegen' selecteren we de statistische functie BINOMIALE.INV en klikken op 'OK' om het venster 'Functieargumenten' te openen;
- in het vak 'Experimenten' voeren we het getal 12 in;
- in het vak 'Kans-gunstig' vullen we de kans op succes in, zijnde 0,5;
- in het vak 'Alfa' voeren we de cumulatieve probabiliteit gelijk aan 0,95 in, omdat we de kritieke waarde willen vinden die de bovenste 5 % afsnijdt van de binomiale kansverdeling. Met andere woorden, de kritieke X-waarde is die waarde waarvan de kans dat een X-waarde daaraan gelijk is of groter is, gelijk is aan α of 5 %;
- we sluiten het venster 'Functieargumenten' en lezen in cel B4 dat de kritieke waarde gelijk is aan 9.

Omdat de kritieke waarde gelijk is aan het aantal waargenomen successen, verwerpen we H_0 en beslissen dat het muntstuk onzuiver is.



B4 =BINOMIALE.INV(12;0,5;0,95)

	A	B	C	D	E	F
1	H_0	p <= 0,5				
2	H_1	p > 0,5				
3	α	0,05				
4	Kritieke waarde	9				
5						

Figuur 48 De binomiale hypothesetoetsfunctie BINOMIALE.INV

2 De Poisson-kansverdeling

2.1 Gebruik

De Poisson-verdeling is een discrete kansverdeling die wordt gebruikt als p of de kans op succes zeer klein is.

De formule voor de berekening van de kans op X successen in een Poisson-verdeling is:

$$P(X) = \frac{\lambda^x \cdot e^{-\lambda}}{x!}$$

waarbij:
X = aantal successen
$\lambda = n \times p = \mu$
e = 2,71828 of het grondtal van de natuurlijke logaritmen

We geven een voorbeeld. Bij de nv TEVERA heeft men vastgesteld dat 3 op de 1 000 geproduceerde dvd-spelers een defect hebben. We vragen ons af wat de kans is dat er op 2 000 toestellen

> 1 toestel defect is;
> 3 of minder toestellen defect zijn.

Voor de berekening van beide kansen becijferen we eerst λ:

$$\begin{aligned} \lambda &= n \times p \\ &= 2\,000 \times 0{,}003 \\ &= 6 \end{aligned}$$

Als we λ en X kennen, kunnen we in de Poisson-kansverdelingstabel (zie bijlage Tabellen) onmiddellijk de kans van X aflezen. Voor X = 1 en $\lambda = 6$ lezen we af dat $P(X = 1) = 0{,}0149$. Voor wat betreft $P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$ vinden we in dezelfde tabel:

$$\begin{aligned} P(X \leq 3) &= 0{,}0025 + 0{,}0149 + 0{,}0446 + 0{,}0892 \\ &= 0{,}1512 \end{aligned}$$

In de praktijk wordt de Poisson-verdeling gebruikt als $p \leq 0{,}1$ en $\lambda \leq 10$.

2.2 POISSON. VERD (POISSON. DIST)

Hierna geven we de gebruikswijze voor de Excelfunctie POISSON. VERD die de waarde van P(X) berekent:

> we selecteren cel B5 in Figuur 49 om het resultaat te tonen;
> we klikken op de f_x-knop en openen het venster 'Functie invoegen', selecteren de functie 'POISSON.VERD', klikken op 'OK' en openen het venster 'Functieargumenten';
> in het X-vak voeren we het aantal defecte eenheden in waarvan we de probabiliteit willen kennen, in ons geval is dat 1;
> in het vak 'Gemiddelden' voeren we λ in of het cijfer 6;
> in het vak cumulatief vullen we 'onwaar' in om de probabiliteit te kennen van juist 1 defecte eenheid.
  We klikken op 'OK' en in cel B5 lezen we het antwoord af: de kans op 1 defecte eenheid is 0,014873.

B5 | =POISSON.VERD(1;6;ONWAAR)

	A	B	C	D	E	F	G
1	p	0,003					
2	n	2000					
3	n*p	6					
4	X	1					
5	P (X = 1)	0,014873					
6							

Figuur 49 De POISSON. VERD-functie: kansberekening van een aantal successen

Als we in het vak 'Cumulatief' WAAR invullen, dan krijgen we als antwoord de cumulatieve probabiliteit.

Als we ons bijvoorbeeld afvragen wat de kans is van $x \leq 3$, dan voeren we in het X-vak het getal 3 in en WAAR in het vak Cumulatief. Het resultaat is dan gelijk aan 0,151204. Figuur 50 toont ons dat resultaat.

B5 | =POISSON.VERD(3;6;WAAR)

	A	B	C	D	E	F	G
1	p	0,003					
2	n	2000					
3	n*p	6					
4	X <=	3					
5	P (X <= 3)	0,151204					
6							

Figuur 50 De POISSON. VERD-functie voor de berekening van een cumulatieve probabiliteit

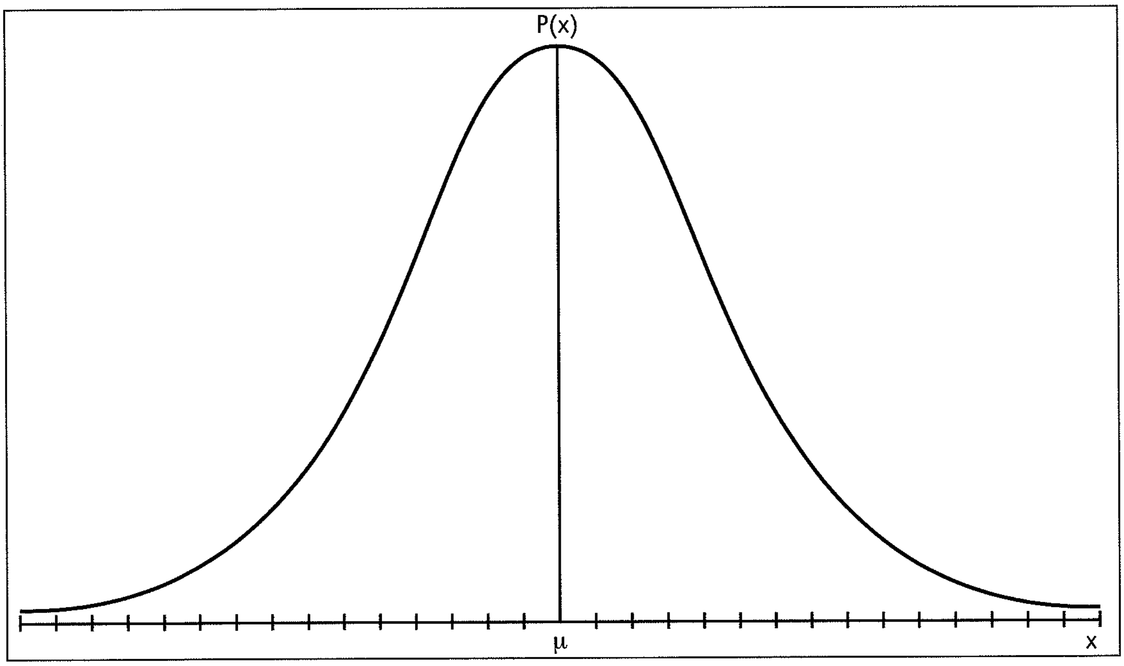
hoofdstuk 12

De normale kansverdeling

1 De grafiek

Als een kenmerk normaal verdeeld is, betekent dat dat de meeste uitkomsten van dat kenmerk rond de gemiddelde uitkomst ervan liggen en dat uitkomsten naarmate ze meer en meer afwijken van de gemiddelde uitkomst, steeds minder worden in aantal.

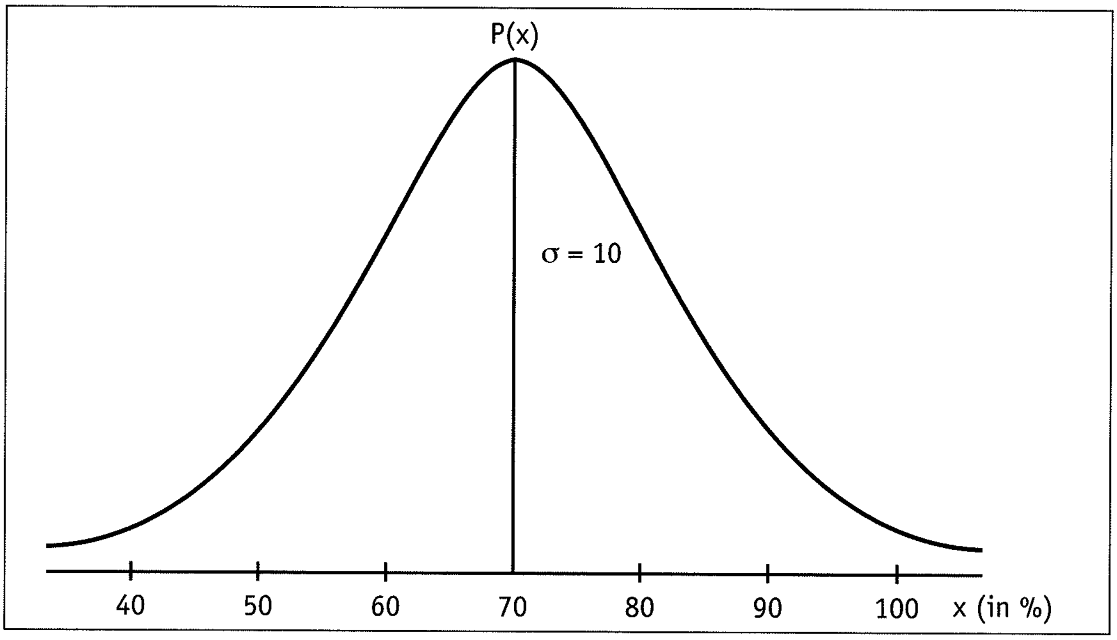
Zo'n verdeling van een kenmerk stelt men grafisch voor door een klokvorm, zoals weergegeven in Figuur 51.



Figuur 51 Grafiek van een normale kansverdeling

Langs de horizontale as vinden we de uitkomsten van het kenmerk X dat we bestuderen. Een verticale lijn vanuit het toppunt van de grafiek naar beneden getrokken, komt uit bij de gemiddelde waarde van X, namelijk μ.
P(x) staat voor de kansdichtheidsfunctie. Het is de functie die ons toelaat om de oppervlakken onder de curve om te zetten in kansen. We beschouwen de kansen van X-waarden die tussen twee X-waarden liggen die overeenstemmen met een bepaald aantal standaardafwijkingen t.o.v. μ.

We verduidelijken dit aan de hand van een voorbeeld waarbij we de uitslagen van een examen waaraan 1 000 mensen hebben deelgenomen, bestuderen. De gemiddelde uitslag is 70 % en de standaardafwijking is 10 %. Omdat we ervan uitgaan dat de uitslagen normaal verdeeld zijn, krijgen we een grafiek zoals in Figuur 52.



Figuur 52 De normale verdeling van de uitslagen van een examen

Tussen 60 en 80, of tussen -1 en +1 standaardafwijking t.o.v. μ, ligt 68 % van de uitslagen.
Tussen 50 en 90, of tussen -2 en +2 standaardafwijkingen t.o.v. μ, ligt 95 % van de uitslagen.
Tussen 40 en 100, of tussen -3 en +3 standaardafwijkingen t.o.v. μ, ligt 99,7 % van de uitslagen.
Bijna de gehele oppervlakte onder de dichtheidskromme ligt dus t.o.v. μ tussen ±3 standaardafwijkingen.
Verder is het interessant dat de oppervlakten tussen ±1, ±2 en ±3 standaardafwijkingen t.o.v. μ respectievelijk met kansen overeenstemmen die voor alle normale verdelingen gelijk zijn.

2 NORM. VERD. N (NORM. DIST)

We zijn natuurlijk niet alleen geïnteresseerd in de kans dat een uitslag ligt tussen twee uitslagen die overeenstemmen met een geheel aantal standaardafwijkingen verwijderd van μ, maar ook in de kans dat een uitslag ligt tussen 65 en 80, tussen 85 en 94 en groter of kleiner is dan 74.

De functie NORM. VERD. N. berekent al deze kansen. Wij geven een uitslag en ook μ en σ van een kansverdeling en NORM. VERD. N berekent de kans dat x kleiner of gelijk is aan de gegeven uitslag.

Bijvoorbeeld de kans dat een uitslag kleiner of gelijk is aan 75 wordt met NORM. VERD. N als volgt gevonden:

- we selecteren cel B2 voor de kansberekening, klikken op f_x, kiezen in het venster 'Functie invoegen' de functie 'NORM. VERD. N', klikken op 'OK' en het venster 'Functieargumenten' verschijnt;
- in het X-vak voeren we de uitslag 75 in, in de vakken 'Gemiddelde' en 'Standaarddev' voeren we respectievelijk 70 en 10 in en in het vak 'Cumulatief' tikken we 'WAAR' in. Door het woord 'WAAR' in te vullen, wordt aan de functie duidelijk gemaakt dat zij de kans moet berekenen dat een uitkomst kleiner of gelijk is aan 75.

In Figuur 53 lezen we af dat $P(x \leq 75) = 0{,}691462$.

B2 | =NORM.VERD.N(75;70;10;WAAR)

	A	B	C	D	E	F
1						
2	P (X<= 75)	0,691462				
3						

Figuur 53 De NORM. VERD. N-functie

Logischerwijze zal de kans dat een uitslag groter is dan 75 gelijk zijn aan 0,308538 (1 – 0,691462).

Als we nu de kans willen berekenen dat een uitslag ligt tussen 63 en 85, dan passen we NORM. VERD. N voor iedere uitslag toe en trekken de berekende kansen van elkaar af. In Figuur 54 wordt de berekening weergegeven.

B4 | =B3-B2

	A	B	C	D
1				
2	P(X<=63)	0,241964		
3	P(X<=85)	0,933193		
4	P(63<X<85)	0,691229		
5				

Figuur 54 De NORM. VERD. N-functie

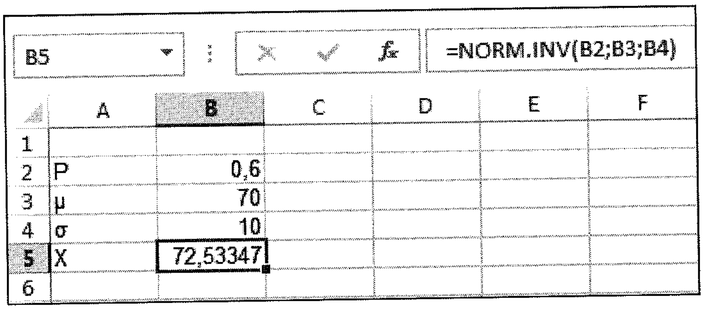
3 NORM. INV. N (NORM. INV)

Met deze functie wordt bij opgave van een kans, een gemiddelde en een standaarddeviatie de uitslag berekend waarvan de kans dat die voorkomt of kleiner is, gelijk is aan de opgegeven probabiliteit.

In Figuur 55 bijvoorbeeld is er opgegeven dat P gelijk is aan 0,60, μ gelijk is aan 70 en σ gelijk is aan 10. De functie NORM. INV. N wordt als volgt gebruikt:

> je selecteert de cel B5 voor de berekening van de uitslag, klikt op f_x en kiest in het venster 'Functie invoegen' voor de functie 'NORM. INV. N'. Je klikt op 'OK' en het venster 'Functieargumenten' wordt geopend;

> in het Kansvak voer je cel B2 in, in het vak 'Gemiddelde' cel B3 en in het vak 'Standaarddev' cel B4. We klikken op 'OK' en de gezochte uitkomst verschijnt in cel B5 en is 72,53347.


B5 =NORM.INV(B2;B3;B4)

	A	B	C	D	E	F
1						
2	P	0,6				
3	μ	70				
4	σ	10				
5	X	72,53347				
6						

Figuur 55 De NORM. INV. N-functie

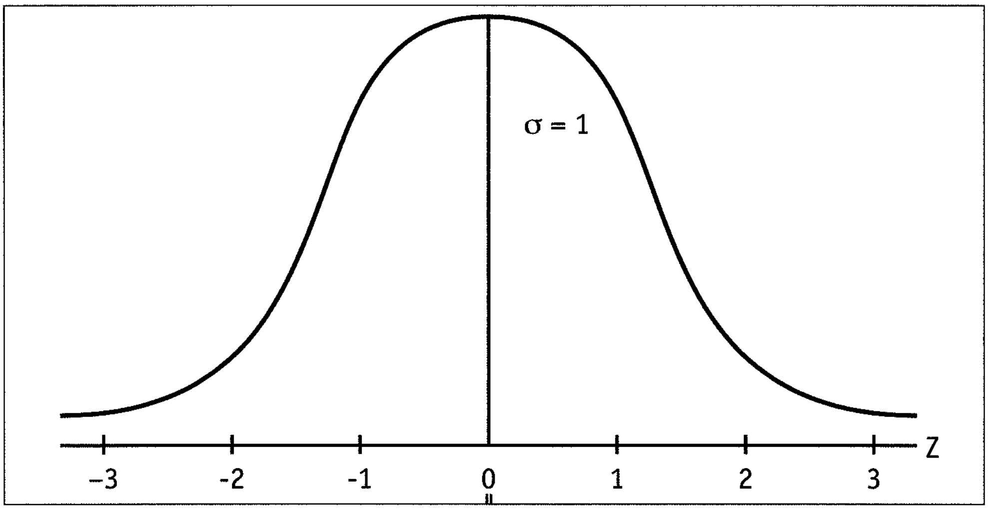
4 De standaardnormale verdeling

Eerder in dit boek hebben we het al gehad over het standaardiseren van X-waarden uit een populatie, het converteren van X-waarden in hun corresponderende Z-waarden met de formule:

$$Z = \frac{x - \mu}{\sigma}$$

Daardoor wordt het mogelijk om waarden uit diverse normale verdelingen met elkaar te vergelijken.

We willen er nu op wijzen dat iedere normale kansverdeling geconverteerd kan worden in een standaardnormale verdeling met μ_z gelijk aan 0 en σ_z gelijk aan 1. De grafiek van de standaardnormale verdeling ziet er dan uit zoals getoond in Figuur 56:



Figuur 56 De standaardnormale kansverdeling

De functies NORM. S. VERD (NORM. S. DIST.) en NORM. S. INV (NORM. S. INV) kunnen we gebruiken voor de Z-verdeling.
Wat is bijvoorbeeld de kans dat Z ligt tussen -2 en +2,5? Om dat te berekenen, passen we de NORM. S. VERD toe en lezen het resultaat af in Figuur 57.

B4 | =B3-B2

	A	B	C	D
1				
2	P(Z<= -2)	0,02275		
3	P(Z<=+2,5)	0,99379		
4	P(-2<Z<+2,5)	0,97104		
5				

Figuur 57 De NORM. S. VERD-functie toegepast op de Z-verdeling

Ten slotte geven we ook nog een voorbeeld met toepassing van de NORM. S. INV-functie op de Z-verdeling.
In Figuur 58 is opgegeven dat P gelijk is aan 0,25, μ gelijk is aan 0 en σ gelijk is aan 1. De te zoeken Z-waarde is gelijk aan -0,67449.

B5 | =NORM.S.INV(0,25)

	A	B	C	D	E	F
1						
2	P	0,25				
3	μ	0				
4	σ	1				
5	Z	-0,67449				
6						

Figuur 58 De NORM. S. INV-functie toegepast op de Z-verdeling

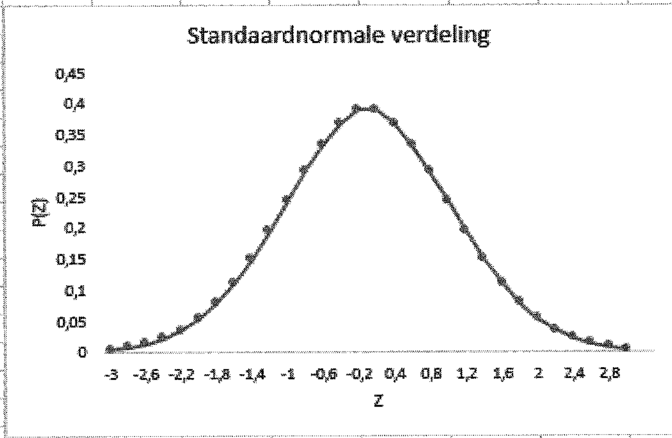
5 De Z-curve

In Figuur 59 voeren we in kolom B van het werkblad een aantal Z-waarden in gaande van -3 tot +3. Vervolgens voeren we in cel C2 de functie 'NORM.S.VERD' in voor Z = -3. Maar omdat we P(Z) of de hoogte van de grafiek van de kansverdeling voor Z = -3 willen berekenen, moeten we in het vak 'Cumulatief' van het venster 'Functieargumenten' de tekst 'ONWAAR' invoeren. Dat is vereist om de kansdichtheidsfunctie P(Z) te kunnen tekenen.

Alle hoogten voor de opgegeven Z-waarden in kolom B vind je vervolgens door de inhoud van cel C2 tot en met cel C32 door te voeren. De grafiek teken je door eerst het bereik C2:C32 te selecteren en dan in het lint het tabblad 'Invoegen' aan te klikken. Kies dan in de keuzelijst 'Lijn- of vlakdiagram invoegen' het subtype 'Lijn met markeringen'. Een Z-curve verschijnt.

Selecteer vervolgens in de groep 'Gegevens', 'Gegevens selecteren'. In het venster 'Gegevensbron selecteren' klik je op 'Bewerken' van 'Horizontale aslabels (categorieën)'. Het venster 'Aslabels' verschijnt. Vul in het vak 'Aslabelbereik' het bereik B2:B32 in. Klik nadien op 'OK' bij 'Aslabels' en 'Gegevensbron selecteren'. Op de horizontale as verschijnen de gewenste Z-waarden.
Om de grafiek een titel te geven, klik je in de groep 'Grafiekindelingen' op de keuzelijst 'Grafiekonderdeel toevoegen' en selecteer je 'Grafiektitel' en 'Boven grafiek'. Als grafiektitel vul je 'Standaardnormale verdeling' in in de formulebalk. Klik daarna op de entertoets.
De volgende stap is de benoeming van de verticale en de horizontale as. Klik in de keuzelijst 'Grafiekonderdeel toevoegen' op 'Astitels' en daarna op 'Primair horizontaal'. Als 'Astitel' voer je 'Z' in. Voor de verticale as selecteer je 'Astitels' opnieuw. Kies 'Primair verticaal' en vul in het 'Astitelvak', 'P(Z)' in. Klik op de linkermuisknop en de ingegeven titel voor de verticale as verschijnt.
Verwijder ten slotte nog de rasterlijnen door in de keuzelijst 'Grafiekonderdelen toevoegen' op 'Rasterlijnen' te klikken en de optie 'Primair groot horizontaal' te selecteren. Figuur 59 geeft het resultaat weer.

	A	B	C
1		Z	P(Z)
2		-3	0,00443
3		-2,8	0,00792
4		-2,6	0,01358
5		-2,4	0,02239
6		-2,2	0,03547
7		-2	0,05399
8		-1,8	0,07895
9		-1,6	0,11092
10		-1,4	0,14973
11		-1,2	0,19419
12		-1	0,24197
13		-0,8	0,28969
14		-0,6	0,33322
15		-0,4	0,36827
16		-0,2	0,39104
17		0	0,39894
18		0,2	0,39104
19		0,4	0,36827
20		0,6	0,33322
21		0,8	0,28969
22		1	0,24197
23		1,2	0,19419
24		1,4	0,14973
25		1,6	0,11092
26		1,8	0,07895
27		2	0,05399
28		2,2	0,03547
29		2,4	0,02239
30		2,6	0,01358
31		2,8	0,00792
32		3	0,00443
33			



Figuur 59 Grafiek van de standaardnormale verdeling

hoofdstuk 13

Intervalschattingen

1 Schatten van parameters

Met behulp van steekproefgrootheden kunnen we populatieparameters schatten. Belangrijk in dat verband is vooraf te bepalen welk betrouwbaarheidsniveau we willen, omdat dit enerzijds het zekerheidsniveau bepaalt dat een populatieparameter binnen een betrouwbaarheidsinterval is begrepen en anderzijds de grootte van het betrouwbaarheidsinterval determineert.

2 Steekproefverdeling van het $\bar{x}$

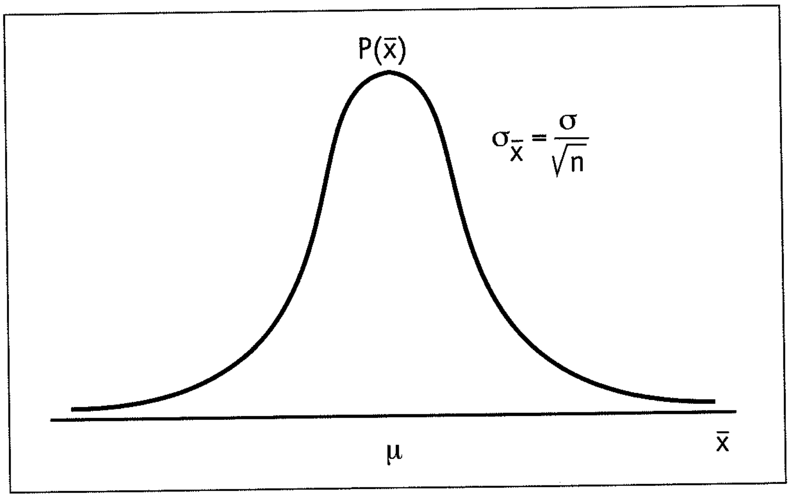
We verkrijgen een steekproefverdeling van het steekproefgemiddelde $\bar{x}$ als we uit een populatie heel veel keren een steekproef van omvang n trekken met teruglegging en daaruit telkens het $\bar{x}$ berekenen, zodat we alle mogelijke waarden van het $\bar{x}$ kennen.

2.1 Vorm

De vorm van de steekproefverdeling van het $\bar{x}$ is die van een normale kansverdeling ongeacht de vorm van de populatieverdeling van zodra $n \geq 30$.

Het gemiddelde van deze verdeling is $\mu_{\bar{x}} = \mu$ en de standaarddeviatie is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

De parameters van deze verdeling vinden we terug in de 'centrale-limietstellingsregel'. Figuur 60 toont ons de klokvorm van de steekproefverdeling van het $\bar{x}$.



Figuur 60 Steekproefverdeling van het $\bar{x}$

3 Betrouwbaarheidsgrenzen van μ

3.1 Klassieke berekeningswijze

Een ziekenhuisdirecteur wil graag weten hoeveel dagen een patiënt gemiddeld in het ziekenhuis verblijft na een operatie en hij wil daarbij een zekerheid van 95 %. Daartoe werd van een steekproef van 50 patiënten het ziekenhuisverblijf na een operatie geregistreerd en het steekproefgemiddelde bedroeg 10 dagen met een standaardafwijking van 5 dagen.

Met deze gegevens kunnen we de grenzen waarbinnen μ met 95 % zekerheid ligt berekenen met behulp van de volgende formule:

$$\mu = \bar{x} \pm Z_{0,025} \times \sigma_{\bar{x}}$$

Eerst berekenen we $\sigma_{\bar{x}}$ of de standaardfout van het steekproefgemiddelde:

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{5}{\sqrt{50}} = 0{,}71$$

Vervolgens lezen we in een kansverdelingstabel voor Z (zie bijlage Tabellen) de Z-waarden af die in beide staarten van de grafiek van de Z-verdeling 2,5 % afsnijden. Die Z-waarden zijn +1,96 en -1,96. Nu kunnen we de formule invullen en we krijgen:

$$\mu = 10 \pm 1{,}96 \times 0{,}71$$
$$\mu = 10 \pm 1{,}39$$

We stellen vast dat de bovengrens 11,39 is en de ondergrens 8,61. We kunnen dus met 95 % zekerheid zeggen dat het gemiddeld aantal dagen dat een patiënt in het ziekenhuis verblijft na een operatie tussen 8,61 dagen en 11,39 dagen ligt.

3.2 VERTROUWELIJKHEID. NORM (CONFIDENCE. NORM)

Deze Excelfunctie berekent de steekproeffout na invoer van het foutenrisico α, op basis waarvan het betrouwbaarheidsniveau wordt berekend, de standaardafwijking s of σ* en de steekproefomvang n.

Ter bepaling van de bovengrens tellen we de steekproeffout ($Z_{0,025} \times \sigma_{\bar{x}}$) op bij het $\bar{x}$. De ondergrens vinden we door de steekproeffout af te trekken van het $\bar{x}$. Hierna volgen de stappen toegepast op het voorbeeld van hierboven:

* Hoewel de toepassing van de Excel-functie aanneemt dat σ steeds bekend is, kunnen wij deze functie ook gebruiken als σ niet bekend is en we enkel s kennen.
$\sigma/\sqrt{n}$, nodig voor de berekening van de steekproeffout, levert geen foutieve steekproeffout op als we σ in $\sigma/\sqrt{n}$ vervangen door s.

- > we voeren de gegevens n, $\bar{x}$, s en het foutenrisico α in, in het bereik B2:B5 van het werkblad in Figuur 61;
- > we selecteren cel B6 voor de berekening van de steekproeffout;
- > we klikken op de f_x-knop en openen het venster 'Functie invoegen';
- > we selecteren de functie 'VERTROUWELIJKHEID. NORM', klikken op 'OK' en openen het dialoogvenster 'Functieargumenten';
- > in het vak 'Alfa' vullen we het foutenrisico in of met andere woorden de kans dat µ niet binnen het berekende betrouwbaarheidsinterval zal liggen, in ons geval is dat 5 %;
- > in het vak 'Standaarddeviatie' vullen we s in en in het vak Grootte de steekproefomvang n;
- > we sluiten het venster 'Functieargumenten' en in cel B6 verschijnt het antwoord, zijnde 1,39 afgerond.

Zoals bij de klassieke berekeningswijze bepalen we de betrouwbaarheidsgrenzen van µ door bij het $\bar{x}$ de steekproeffout respectievelijk af te trekken en op te tellen.

B6 | =BETROUWBAARHEID(0,05;B4;B2)

	A	B	C	D	E	F
1						
2	n	50				
3	steekproefgemiddelde	10				
4	s	5				
5	α	5%				
6	steekproeffout	1,386				
7	bovengrens	11,386				
8	ondergrens	8,614				
9						

Figuur 61 De VERTROUWELIJKHEID. NORM-functie

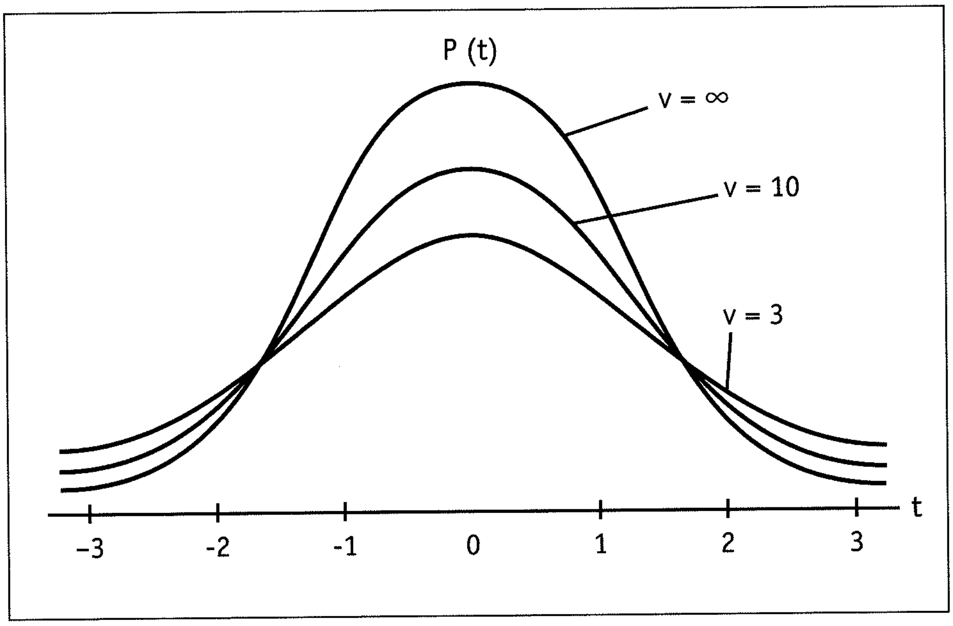
4 t-verdeling

4.1 Klassieke berekeningswijze van de betrouwbaarheidsgrenzen van µ

Als het aantal steekproefelementen kleiner is dan 30, dan is de steekproefverdeling van het $\bar{x}$ een t-kansverdeling. Deze verdeling is ook klokvormig* en de spreiding van de verdeling is afhankelijk van het aantal vrijheidsgraden v of het aantal onafhankelijke steekproefelementen. Dit betekent dat de toe te passen t-kansverdeling afhankelijk is van het aantal vrijheidsgraden in de steekproef.

* Op voorwaarde dat de populatieverdeling van X normaal is.

Voor de t-verdeling van het $\bar{x}$ is het aantal vrijheidsgraden gelijk aan n – 1. Onderstaande Figuur 62 toont een aantal t-verdelingen afhankelijk van de omvang van n en illustreert dat als v = ∞ de t-verdeling gelijk is aan de normale verdeling.



Figuur 62 t-verdelingen

We hernemen het voorgaande voorbeeld en nemen aan dat het aantal steekproefelementen nu gelijk is aan 20.
De formule die ons toelaat om het 95 %-betrouwbaarheidsinterval te berekenen is:

$$\mu = \bar{x} \pm t_{0,025} \times \frac{s}{\sqrt{n}}$$

In een kansverdelingstabel voor t-waarden (zie bijlage Tabellen) lezen we af dat voor v = 20 – 1 de t-waarden +2,09 en -2,09 elk in beide staarten van de betreffende t-verdeling 2,5 % afsnijden.

Als we de formule voor de berekening van ons betrouwbaarheidsinterval invullen, dan krijgen we:

$$\mu = 10 \pm 2{,}09 \times \frac{5}{\sqrt{20}}$$

$$\mu = 10 \pm 2{,}34$$

De bovengrens is 12,34 en de ondergrens 7,66.
Met 95 % zekerheid kunnen we stellen dat μ of het gemiddeld aantal ziektedagen dat een patiënt in het ziekenhuis doorbrengt na een operatie ligt tussen 7,66 en 12,34 dagen.

We stellen vast – zoals verwacht – dat het interval groter is geworden ten gevolge van het kleinere aantal steekproefelementen.

4.2 VERTROUWELIJKHEID. T (CONFIDENCE. T)

Deze functie laat toe de steekproeffout te vinden voor v = n – 1 en een vooraf opgegeven foutenrisico dat verband houdt met een tweezijdige t-verdeling.

De functie toegepast op het voorbeeld, werkt als volgt:

> we voeren de vereiste gegevens in in het werkblad van Figuur 63: n, $\bar{x}$, s* en het foutenrisico;
> we selecteren cel B6 voor de berekening van de steekproeffout;
> we klikken op de f_x-knop, openen het venster 'Functie invoegen', selecteren 'VERTROUWELIJKHEID. T' en openen het venster 'Functieargumenten';
> in het vak 'ALFA' vullen we het foutenrisico in, zijnde 5 % en in het vak 'Standaarddev' vullen we s in;
> in het vak 'grootte' vullen we het aantal steekproefelementen in, zijnde 20;
> we sluiten het venster 'Functieargumenten' en in cel B6 lezen we de gezochte steekproeffout: 2,34 afgerond.

De betrouwbaarheidsgrenzen zijn dus:

bovengrens: 10 + 2,34 = 12,34;

ondergrens: 10 - 2,34 = 7,66.

B6 | =VERTROUWELIJKHEID.T(0,05;5;20)

	A	B	C	D	E	F
1						
2	n	20				
3	$\bar{X}$	10				
4	s	5				
5	foutenrisico α	0,05				
6	Steekproeffout	2,34				
7						

Figuur 63 De VERTROUWELIJKHEID. T-functie

* Deze functie veronderstelt dat σ steeds bekend is. In de praktijk is dit meestal niet het geval en mogen we desgevallend σ vervangen door s.

hoofdstuk 14

De hypothesetest

1 Hypothesen

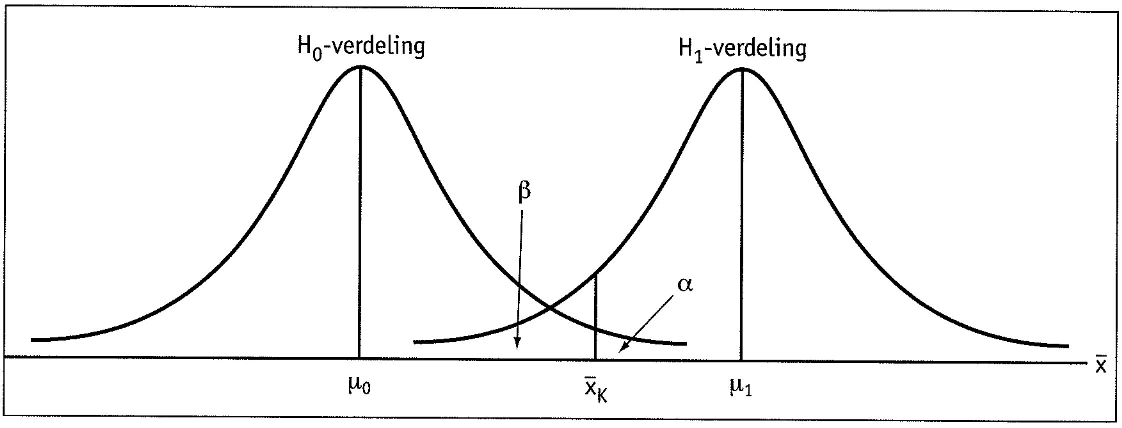
Een hypothese is een veronderstelling die we maken over een populatie en die we kunnen onderzoeken op basis van een steekproefstudie.

Voor het uitvoeren van een hypothesetest, formuleren we twee hypothesen:

- de **nulhypothese** of H_0 stelt dat de steekproefresultaten geen bewijsmateriaal opleveren voor het verwerpen van een bestaande norm;
- de **alternatieve hypothese** of H_1 stelt dat de steekproefresultaten wel bewijsmateriaal opleveren voor het verwerpen van een bestaande norm.

Bij de beslissing die we nemen om H_0 te aanvaarden of te verwerpen, kunnen we twee types van fouten maken.
Een eerste fout is het ten onrechte verwerpen van H_0. De kans daarop wordt vooraf bepaald door α.
Een tweede type van fout is het ten onrechte aanvaarden van H_0 waardoor we ten onrechte H_1 verwerpen. De kans daarop wordt voorgesteld door β.

Het voorgaande tonen we in Figuur 64, waar we geconfronteerd worden met de vraag of een $\bar{x}$ berekend uit een steekproef ($n \geq 30$) al of niet bewijs levert tegen μ_0.



Figuur 64 De steekproefverdelingen van H_0 en H_1 met aanduiding van α en β

De kritieke waarde $\bar{x}_K$ zal bepalen tot welke steekproefverdeling ons $\bar{x}$ zal behoren. Om $\bar{x}_K$ te kunnen berekenen, moeten we eerst α bepalen. Meestal wordt α gelijkgesteld aan een probabiliteit van 5 %.

Als het berekende $\bar{x}$ groter is dan $\bar{x}_K$, dan verwerpen we H_0; in het omgekeerde geval aanvaarden we H_0.
Figuur 64 maakt duidelijk dat er een kans α bestaat dat we H_0 ten onrechte verwerpen en een kans β dat we H_1 ten onrechte verwerpen.

2 Traditionele hypothesetoets

We nemen het voorbeeld van een onderneming die autobanden produceert, waarvan de gemiddelde levensduur 50 000 km is en de standaarddeviatie gelijk is aan 6 250 km. De ingenieurs van het bedrijf hebben met behulp van een nieuwe grondstof een autoband ontwikkeld die volgens hen langer gebruikt kan worden dan de bestaande autoband. De nieuwe autoband wordt getest op basis van een steekproef van 100 autobanden en de gemiddelde levensduur is 51 250 km.

We formuleren de eenzijdige hypothesetoets:

H_0: $\mu_0 = 50\ 000$

H_1: $\mu_1 > 50\ 000$

$\alpha = 5\ \%$

We berekenen eerst de Z-waarde van het gevonden $\bar{x}$:

$$Z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{51\ 250 - 50\ 000}{\frac{6\ 250}{\sqrt{100}}} = \frac{1\ 250}{625} = 2$$

Ten tweede berekenen we op basis van $\alpha = 5\ \%$ de kritieke Z-waarde die 5 % van het oppervlak van de standaardnormale kansverdeling afsnijdt in de rechterstaart.
Met de functie NORM. S. INV in Figuur 65 vinden we een Z-waarde gelijk aan +1,64.

B5 =NORM.S.INV(0,95)

	A	B	C	D	E	F
1						
2	P	0,95				
3	μ	0				
4	σ	1				
5	Z	1,64				
6						

Figuur 65 Berekening van de kritieke Z-waarde

Omdat Z > 1,64 kunnen we H_0 verwerpen met een foutenrisico van 5 %. De kans dat Z gelijk is aan 2 als μ_0 waar is, is kleiner dan 5 % en dat levert een bewijs op tegen H_0.

Een hypothese kan ook tweezijdig zijn en dan stelt H_1 in ons voorbeeld dat μ_1 verschillend is van μ_0. Dat brengt ons dan tot de volgende formulering:

H_0: $\mu_0 = 50\ 000$

H_1: $\mu_1 \neq 50\ 000$

$\alpha = 5$ %

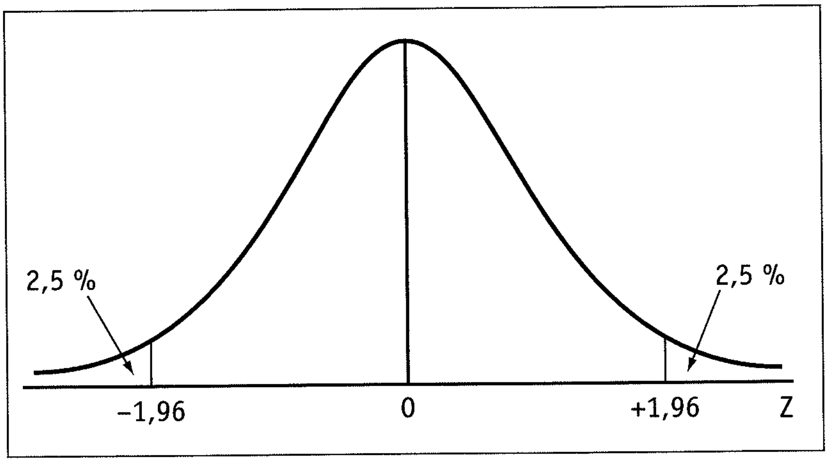
Nu moeten we het tweezijdige verwerpingsgebied berekenen dat in iedere staart van de standaardnormale kansverdeling 2,5 % afsnijdt. Met de functie NORM. S. INV in Figuur 66 vinden we een Z-waarde gelijk aan +1,96 die 2,5 % in de rechterstaart afsnijdt. Dat betekent dat -1,96 in de linkerstaart ook 2,5 % zal afsnijden.

B5 =NORM.S.INV(0,975)

	A	B	C	D	E	F
1						
2	P	0,975				
3	μ	0				
4	σ	1				
5	Z	1,96				
6						

Figuur 66 Berekening van de kritieke Z-waarde

Omdat de gevonden Z-waarde gelijk is aan 2 en dus in het verwerpingsgebied ligt zoals aangegeven in Figuur 67, beslissen we opnieuw om H_0 te verwerpen.



Figuur 67 Tweezijdig verwerpingsgebied als $\alpha = 5$ %

3 Z. TEST (Z. TEST)

De Z. TEST wordt gebruikt voor het testen van hypothesen waarbij een beroep kan worden gedaan op de Z-verdeling.

De invoer bestaat uit steekproefgegevens, de formulering van H_0 en de waarde van σ. Met deze gegevens berekent de functie onmiddellijk de kans dat het $\bar{x}$ groter is dan het waargenomen $\bar{x}$ als H_0 waar is.

We veronderstellen dat het lichaamsgewicht van arbeiders lager is dan het gemiddelde lichaamsgewicht van de werknemers in een onderneming. Om deze stelling te testen, registreerden we van 20 arbeiders het lichaamsgewicht waarvan het $\bar{x}$ gelijk is aan 71,7 kg. We weten ook dat μ_0 gelijk is aan 78 kg, σ gelijk is aan 11 kg en α gelijk is aan 0,05. Zo kunnen we de Z. TEST uitvoeren:

> in Figuur 68 voer je in het bereik B3:B22 de steekproefgegevens in en selecteer je cel B23 voor de gezochte kans van het $\bar{x}$ als H_0 waar is;

> je klikt op de f_x-knop, opent het venster 'Functie invoegen' en kiest daarin de Z. TEST;

> in het venster 'Functieargumenten' voer je in het matrixvak het gegevensbereik B3:B22 in;

> in het X-vak vul je de te testen waarde van μ_0 in, in dit geval 78, en in het Sigma-vak voeren we σ in, zijnde 11;

> we sluiten het venster en in cel B23 lezen we het resultaat: 0,995. Dat betekent dat de kans dat het $\bar{x}$ gelijk is aan het geobserveerde $\bar{x}$ of kleiner is, gelijk is aan 0,005. Omdat deze kans kleiner is dan α, verwerpen we H_0 en besluiten we dat het lichaamsgewicht van de arbeiders lager is dan het gemiddelde gewicht van alle werknemers.

B23 =Z.TOETS(B3:B22;78;11)

	A	B	C	D	E	F
1						
2		Gewicht				
3		67				
4		63				
5		72				
6		80				
7		69				
8		74				
9		81				
10		62				
11		75				
12		69				
13		70				
14		61				
15		85				
16		79				
17		67				
18		73				
19		76				
20		68				
21		70				
22		73				
23	Z.TEST	0,995				
24	P	0,005				
25						

Figuur 68 De eenzijdige Z. TEST

Als we van deze eenzijdige toets een tweezijdige toets maken of als we stellen dat het gemiddelde gewicht van de arbeiders verschillend is van het gemiddelde gewicht van alle personeelsleden, dan is de kans dat het $\bar{x}$ langs weerszijden op eenzelfde of grotere afstand van μ_0 ligt, gelijk aan twee keer 0,005 of 0,010. Omdat deze kans kleiner is dan α, verwerpen we opnieuw H_0.

4 T. DIST (T. DIST)

Als σ onbekend is en $n < 30$, dan gebruiken we de statistische grootheid t, met als formule:

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

Ter illustratie nemen we het volgende voorbeeld: een steekproef van 16 uitslagen voor het vak economie gaf een $\bar{x} = 69$ op 100 punten en een $s = 14$. Er wordt gesteld dat μ gelijk is aan 60 op 100 punten. We vragen ons nu af of we deze H_0-hypothese kunnen aanhouden bij een $\alpha = 5$ %.

Als t-waarde vinden we:

$$t = \frac{69 - 60}{\frac{14}{\sqrt{16}}} = \frac{9}{3{,}8} = 2{,}57$$

Met gebruik van de T. DIST-functie kunnen we onderzoeken of we H_0 al dan niet moeten verwerpen:

> selecteer in het werkblad van Figuur 69 cel B3 om het resultaat te berekenen;
> klik op de f_x-knop en selecteer in het venster 'Functie invoegen' de T. DIST;
> in het venster 'Functieargumenten' voer je in het X-vak de berekende waarde 2,57 voor t in, in het vak 'Vrijheidsgraden' vul je het getal 15 in en ten slotte in het vak 'Cumulatief' typen we 'WAAR' in. Daardoor weet de functie dat zij de kans moet berekenen dat t kleiner of gelijk is aan 2,57.

Omdat de kans dat t groter is dan de berekende t-waarde gelijk is aan slechts 0,0107 (1 – 0,9893) moeten we natuurlijk H_0 verwerpen en besluiten dat μ groter is.

B3 | =T.DIST(2,57;15;WAAR)

	A	B	C	D	E	F
1						
2						
3	P	0,9893				
4						

Figuur 69 De T. DIST voor een eenzijdige hypothesetoets

5 Variantietoets met χ^2

Met de chikwadraatgrootheid kunnen we hypothesen testen in verband met varianties die een normale verdeling hebben.

Afhankelijk van het aantal vrijheidsgraden moeten we een bepaalde χ^2-verdeling gebruiken. De formule voor χ^2 is:

$$\chi^2 = \frac{(n - 1) \times s^2}{\sigma^2}$$

We veronderstellen dat ringen geproduceerd door een machine voor hun diameter hoogstens een standaardafwijking van 0,5 mm mogen hebben. Na meting van 20 ringen, vinden we een steekproefstandaardafwijking van 0,65 mm. De hypothesen zijn:

H_0: $\sigma^2 \leq 0{,}25$

H_1: $\sigma^2 > 0{,}25$

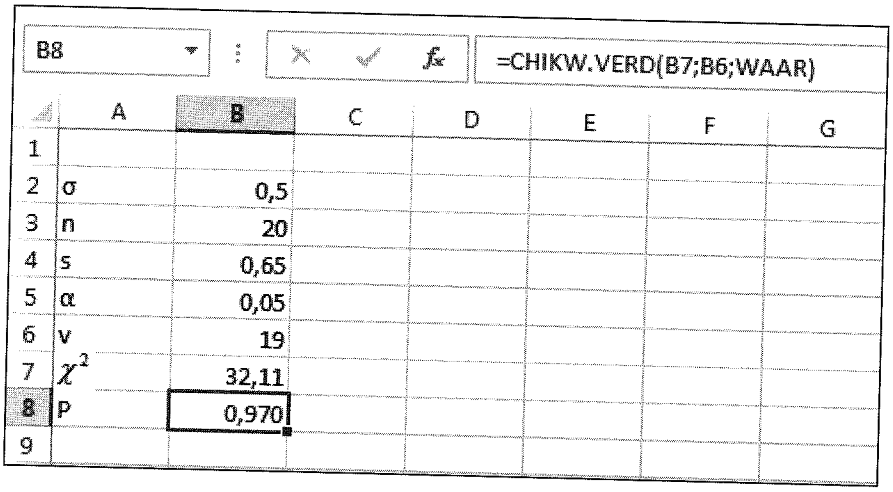
$\alpha = 0{,}05$

We berekenen:

$$\chi^2 = \frac{(20 - 1) \times (0{,}65)^2}{(0{,}5)^2} = \frac{19 \times 0{,}423}{0{,}25} = 32{,}11$$

Met behulp van de functie CHIKW. VERD (CHISQ. DIST) kunnen we de kans berekenen dat χ^2 kleiner of gelijk is aan 32,11 als H_0 waar is. Als deze kans groter is dan 0,95 $(1 - \alpha)$, dan verwerpen we H_0.
We tonen hoe de functie werkt:

> we selecteren in een werkblad cel B8 voor het resultaat van de functie;
> we klikken op de f_x-knop om het venster 'Functie invoegen' te openen en kiezen 'CHIKW. VERD';
> in het venster 'Functieargumenten' voeren we in het X-vak de berekende χ^2-waarde in en in het vak 'Vrijheidsgraden' het aantal vrijheidsgraden, zijnde 19 (20 – 1);
> in het vak 'Cumulatief' typen we 'WAAR' in;
> we sluiten het venster 'Functieargumenten' en we lezen in cel B8, zoals getoond in Figuur 70, de eenzijdige kans dat χ^2 gelijk is aan 32,11 of kleiner is dan 32,11 als H_0 waar is. Omdat deze kans groter is dan 0,95 $(1 - \alpha)$, verwerpen we H_0 en aanvaarden dat $\sigma^2 > 0{,}25$ en dat de machine niet voldoende accuraat werkt.



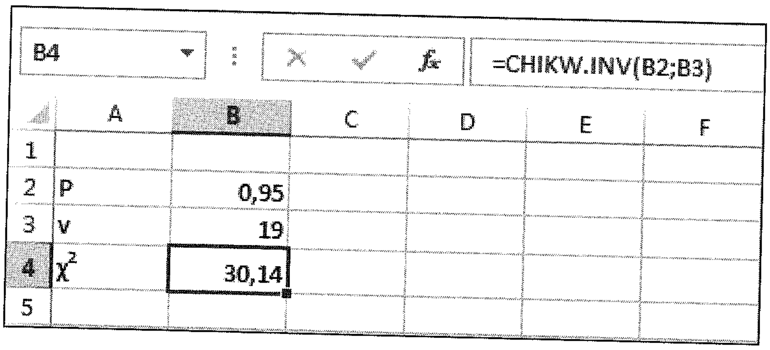
Figuur 70 De CHIKW. VERD-functie

De functie CHIKW. INV (CHISQ. INV) berekent de χ^2-waarde waarvan de kans dat ze voorkomt kleiner is of gelijk aan de opgegeven probabiliteit, daarbij rekening houdende met het aantal vrijheidsgraden v. Dit betekent in ons voorbeeld de berekening van de kritieke χ^2-waarde om uit te maken of we H_0 al dan niet zullen verwerpen.

De functie werkt als volgt:

> we selecteren in Figuur 71 de cel B4 voor de te berekenen χ^2-waarde;
> we klikken op de f_x-knop en kiezen in het venster 'Functie invoegen' de functie CHIKW. INV;
> in het venster Functieargumenten voeren we in het vak 'Kans' de kans 0,95 in en in het vak 'Vrijheidsgraden' zetten we 19. We sluiten het venster en lezen in cel B4 dat de kritieke χ^2-waarde gelijk is aan 30,14. Deze waarde snijdt de bovenste 5 % af van het oppervlak van deze specifieke chikwadraatverdeling.

Aangezien de berekende waarde in ons voorbeeld 32,11 was, levert de CHIKW. INV-functie eveneens een bewijs tegen H_0.



Figuur 71 De CHIKW. INV-functie

6 CHIKW. TEST (CHISQ. TEST) voor onafhankelijkheid

De functie CHIKW. TEST wordt gebruikt om te onderzoeken of er een verband bestaat tussen twee variabelen. H_0 stelt dan dat er onafhankelijkheid is tussen de twee te onderzoeken variabelen en H_1 verwerpt uiteraard H_0.

Als voorbeeld nemen we 300 proefpersonen waarvan we onderzoeken of de keuze van hun krantentitel A, B of C afhangt van hun studieniveau. De steekproefresultaten worden getoond in het werkblad van Figuur 72.
We vragen ons nu af of de gegevens aantonen dat de keuze van een krantentitel onafhankelijk is van het studieniveau. Om dat na te gaan, formuleren we de volgende hypothesetest:

H_0: keuze van een krantentitel is onafhankelijk van het studieniveau

H_1: H_0 wordt verworpen

$\alpha = 5$ %

Om onze hypothesen te toetsen, moeten we de verwachte frequenties berekenen van alle combinaties van X en Y in de veronderstelling dat beide onafhankelijk zijn. Daarvoor moeten we de twee volgende stappen uitvoeren:

a Van alle combinaties X en Y de geschatte kans berekenen met de kansformule $P(X,Y) = P(X) \times P(Y)$. Dat gebeurt aan de hand van de geschatte kansverdelingen van P(X) en P(Y):

				P(X)
	0,38 × 0,35 = 0,13	0,31 × 0,35 = 0,11	0,31 × 0,35 = 0,11	0,35
	0,38 × 0,30 = 0,12	0,31 × 0,30 = 0,09	0,31 × 0,30 = 0,09	0,30
	0,38 × 0,22 = 0,08	0,31 × 0,22 = 0,07	0,31 × 0,30 = 0,07	0,22
	0,38 × 0,13 = 0,05	0,31 × 0,13 = 0,04	0,31 × 0,13 = 0,04	0,13
P(Y)	0,38	0,31	0,31	

b Met de berekende geschatte kansen het verwachte aantal personen per combinatie bepalen door iedere kans te vermenigvuldigen met het totale aantal proefpersonen. Het resultaat daarvan vinden we ook in het werkblad van Figuur 72.

Nadat we de verwachte frequenties hebben berekend, kunnen we deze vergelijken met de waargenomen frequenties, met behulp van de volgende formule:

$$\chi^2 = \Sigma \frac{(0 - E)^2}{E}$$

Met de uitkomst van χ^2 testen we in de overeenstemmende chikwadraatverdeling of H_0 al of niet aanvaard kan worden. In dit voorbeeld heeft de χ^2-verdeling 6 vrijheidsgraden of v = (4 – 1) × (3 – 1). De uitvoering van de test doen we met de CHIKW. TEST-functie:

> selecteer cel C22 in Figuur 72;

> je klikt op de f_x-knop en in het venster 'Functie invoegen' kies je de functie 'CHIKW. TEST' en klikt op 'OK';

> in het dialoogvenster 'Functieargumenten' breng je in het vak 'Waarnemingen' het gegevensbereik C4:E7 in en in het vak 'Verwacht' het bereik C15:E18 en je sluit het venster;

> in cel C22 vinden we voor de berekende uitkomst van χ^2 een kans van 7,36208E – 06 of 0,000007.

Omdat de kans van H_0 kleiner is dan 0,05, verwerpen we H_0 en besluiten dat de keuze van een krantenartikel afhankelijk is van het studieniveau.

C22 | =CHIKW.TEST(C4:E7;C15:E18)

	A	B	C	D	E	F	G	H
1								
2		Tabel Studieniveau en Krantentitels (waargenomen frequenties)						
3			Krant A	Krant B	Krant C	Totaal	P(X)	
4		M.O.	60	30	15	105	0,35	
5		H.O. 1 cyclus	30	30	30	90	0,30	
6		H.O. 2 cycli	15	20	30	65	0,22	
7		U.O.	10	12	18	40	0,13	
8		Totaal	115	92	93	300		
9		P(Y)	0,38	0,31	0,31			
10								
11								
12								
13		Tabel Studieniveau en Krantentitels (verwachte frequenties)						
14			Krant A	Krant B	Krant C	Totaal		
15		M.O.	40	32	33	105		
16		H.O. 1 cyclus	35	28	28	91		
17		H.O. 2 cycli	25	20	20	65		
18		U.O.	15	12	12	39		
19		Totaal	115	92	93	300		
20								
21								
22		P	7,36208E-06					
23								

Figuur 72 De CHIKW. TEST-functie

hoofdstuk 15

Tweesteekproeven-hypothesetoetsen

1 Testen van het verschil van twee steekproefgemiddelden als de populatievarianties gekend zijn

Voor het testen van hypothesen waarbij een beroep gedaan wordt op het verschil van twee steekproefgemiddelden, hebben we een nieuwe steekproefverdeling nodig, namelijk de steekproefverdeling van het verschil tussen gemiddelden die bestaat uit alle mogelijke verschillen tussen paren van steekproefgemiddelden.

Omdat beide steekproeven per paar worden getrokken uit een verschillende verzameling, zegt men dat de steekproeven onafhankelijk zijn.

De grafiek van deze verdeling is normaal of dus klokvormig, op voorwaarde dat de steekproeven groot zijn of, als ze klein zijn, dat hun populaties normaal verdeeld zijn.

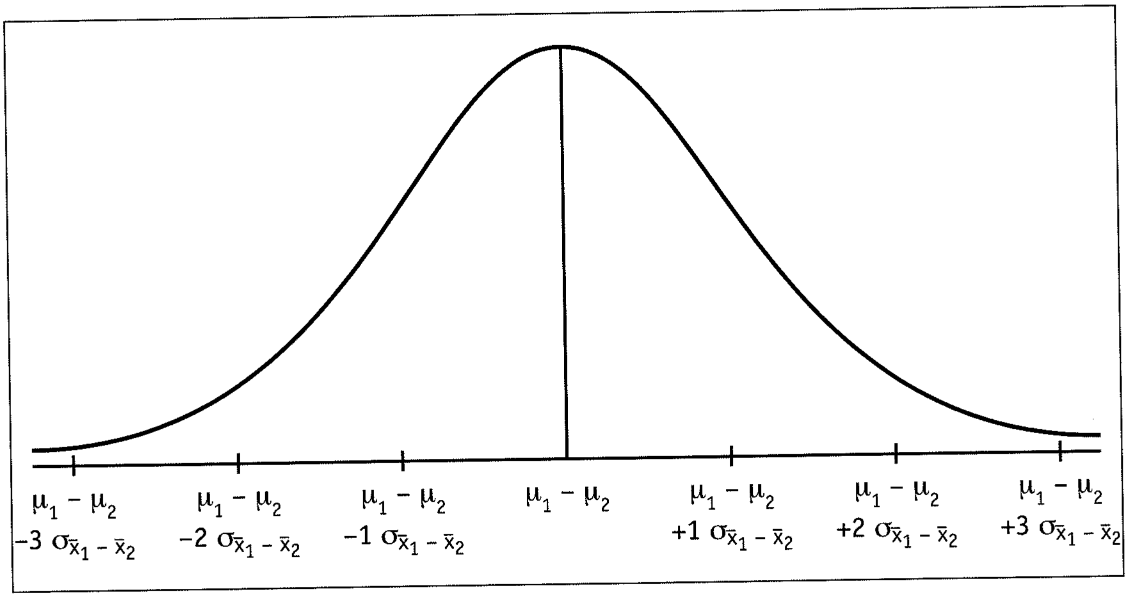
Het gemiddelde van deze steekproefverdeling is:

$$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$$

en de standaarddeviatie is:

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_{x1}^2}{n_1} + \frac{\sigma_{x2}^2}{n_2}}$$

De onderstaande Figuur 73 toont deze steekproefverdeling met haar parameters.



Figuur 73 Steekproefverdeling van het verschil tussen twee steekproefgemiddelden

Als de verdeling normaal is en de populatievarianties σ_1^2 en σ_2^2 gekend zijn, dan kunnen we de Z-verdeling gebruiken. De formule voor de berekening van Z is dan:

$$Z = \frac{\bar{x}_1 - \bar{x}_2 - \mu_1 - \mu_2}{\sigma_{\bar{x}_1 - \bar{x}_2}}$$

We geven het volgende voorbeeld. Een onderneming heeft net een promotiecampagne gedaan voor haar nieuwe frisdrankproduct 'Biotan'. Om het succes van de campagne te beoordelen, controleerde de reclameafdeling 20 klanten vóór de campagne en 20 klanten na de promotieactie. Het gemiddelde verbruik voor de campagne was 1,80 l en na de campagne was dat 2,20 l.
Verder nemen we aan dat de populatiestandaarddeviatie vóór en na de campagne identiek is en gelijk is aan 1 l.
Als α = 5 % kan men dan besluiten dat de promotieactie de vraag naar Biotan heeft doen toenemen?

De hypothesen zijn:

H_0: $\mu_1 - \mu_2 = 0$

H_1: $\mu_1 - \mu_2 > 0$

We berekenen Z voor het geval H_0 waar is:

$$Z = \frac{2{,}2 - 1{,}8 - \mu_1 - \mu_2}{\sqrt{\frac{(1)^2}{20} + \frac{(1)^2}{20}}} = \frac{0{,}4 - 0}{\sqrt{0{,}05 + 0{,}05}} = \frac{0{,}4}{0{,}3162} = +1{,}26$$

Met de Excelfunctie NORM. S. INV vinden we voor α = 5 % een kritieke Z-waarde van 1,64, zijnde de waarde die de bovenste 5 % van het oppervlak afsnijdt onder de curve van de standaardnormale kansverdeling. Omdat onze berekende Z-waarde kleiner is dan de kritieke Z-waarde of 1,26 < 1,64, verwerpen we H_0 niet.

2 Z-toets: twee steekproeven voor gemiddelden (Z-test: Two sample for means)

Excel beschikt over een gegevensanalysefunctie waarmee we een toets, zoals we die hierboven hebben getoond, eenvoudig kunnen uitvoeren. Deze toets is de Z-toets: twee steekproeven voor gemiddelden. De functie wordt als volgt gebruikt:

- in het werkblad van Figuur 74 voeren we de gegevens van twee steekproeven in, in twee verschillende gegevensmatrices: de gegevens van het verbruik vóór de promotiecampagne in kolom C en die van na de promotiecampagne in kolom B;
- in het lint selecteren we het tabblad 'Gegevens'. Vervolgens klikken we op de knop 'Gegevensanalyse' waarmee we het dialoogvenster 'Gegevensanalyse' openen. In dat venster kiezen we de functie 'Z-toets': twee steekproeven voor gemiddelden' en klikken op 'OK' om de functie te openen;
- in het invoervak 'Variabelenbereik 1' voeren we het gegevensbereik B1:B21 in;
- in het invoervak 'Variabelenbereik 2' plaatsen we het gegevensbereik C1:C21;
- in het vak 'Schatting verschil tussen gemiddelden' tikken we het verschil tussen μ_1 en μ_2 in volgens H_0, zijnde 0;
- in het vak 'Variantie van variabele 1' voeren we de variantie voor de eerste steekproef in, dat is $(1)^2$ of 1;
- in het vak 'Variantie van variabele 2' komt de variantie voor de tweede steekproef: 1;
- als de gegevensbereiken kolomtitels bevatten, dan selecteren we de optie 'Labels'. Dat is hier het geval, dus duiden we de optie aan;
- in het 'Alfa-vak' voeren we een foutenrisico in van 5 %;
- bij de uitvoeropties selecteren we 'Nieuw werkblad' om de resultaten op een nieuw blad in het werkblad te tonen;
- we klikken op 'OK' en het resultaat zien we in Figuur 75.

	A	B	C	D
1		Verbruik na campagne	Verbruik voor campagne	
2		2,70	1,70	
3		3,60	2,90	
4		0,70	0,80	
5		2,30	1,60	
6		2,00	1,80	
7		2,90	1,80	
8		3,00	3,50	
9		3,00	2,00	
10		2,00	0,80	
11		2,20	2,60	
12		1,80	1,30	
13		0,70	0,90	
14		1,90	1,60	
15		2,10	2,40	
16		1,90	0,80	
17		2,10	1,90	
18		1,90	2,00	
19		3,80	3,10	
20		0,50	0,60	
21		2,90	1,90	
22				

Figuur 74 Gegevens van twee steekproeven

	A	B	C	D
1	Z-toets: twee steekproeven voor gemiddelden			
2				
3		*Verbruik na campagne*	*Verbruik voor campagne*	
4	Gemiddelde	2,2	1,8	
5	Bekende variantie	1	1	
6	Waarnemingen	20	20	
7	Schatting van verschil tussen gemiddelden	0		
8	z	1,264911064		
9	P(Z<=z) eenzijdig	0,102951671		
10	Kritiek gebied voor Z-toets: eenzijdig	1,644853		
11	P(Z<=z) tweezijdig	0,205903342		
12	Kritiek gebied voor Z-toets: tweezijdig	1,959961082		
13				

Figuur 75 Z-toets: twee steekproeven voor gemiddelden

Rijen 4, 5 en 7 bevatten gegevens die we in de dialoogvakken hebben ingevoerd. In rij 6 vinden we het aantal steekproefgegevens van beide steekproeven.

In cel B8 lezen we de Z-waarde af voor het verschil van de twee steekproefgemiddelden als H_0 waar is.

De kritieke Z-waarde staat in cel B10 en in cel B12 lezen we de kritieke waarde af voor een tweezijdige toets.

Cel B9 geeft het oppervlak weer dat onze berekende Z-waarde afsnijdt in één staart van de standaardnormale kansverdeling en cel B11 geeft het oppervlak dat wordt afgesneden door de positieve en negatieve waarde van de berekende Z-waarde in beide staarten van de kansverdeling.

Uiteraard zullen we opnieuw H_0 aanvaarden omdat $P(Z \geq 1{,}26) = 0{,}10$ of groter is dan α.

3 Testen van het verschil van twee steekproefgemiddelden als de populatievarianties niet gekend zijn

In dit geval moeten we de t-verdeling gebruiken en dus een t-test uitvoeren. Ook moeten we rekening houden met de mogelijkheid dat de ongekende populatievarianties ofwel identiek zijn ofwel ongelijk zijn.

3.1 De ongekende populatievarianties zijn gelijk

In deze situatie zijn beide steekproefvarianties een schatter van σ^2. Daarom voegen we de steekproefvarianties samen of combineren we ze om σ^2 te schatten. Dat gebeurt door de berekening van s_G^2, de gecombineerde schatter van σ^2:

$$s_G^2 = \frac{(n_1 - 1) \times s_1^2 + (n_2 - 1) \times s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

De formule voor t wordt dan:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_G \times \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Ter verduidelijking geven we het volgende voorbeeld. In onderneming A wordt een steekproef van 10 wedden getrokken: € 1 350, € 1 700, € 1 850, € 2 000, € 1 800, € 1 400, € 1 475, € 1 625, € 1 500 en € 1 390. In een tweede onderneming B trekt men eveneens een steekproef van 10 wedden: € 1 250, € 1 750, € 1 600, € 2 100, € 1 520, € 1 500, € 1 375, € 1 725, € 1 400 en € 1 460. We willen nu weten of er tussen de weddegemiddelden van de bedienden van beide ondernemingen een verschil bestaat.
De hypothesen voor deze tweezijdige toets zijn dan:

H_0: $\mu_1 - \mu_2 = 0$

H_1: $\mu_1 - \mu_2 \neq 0$

$\alpha = 0{,}05$

In Figuur 76 tikken we in kolommen A en D onze steekproefgegevens in en berekenen met de functie GEMIDDELDE van beide steekproeven het $\bar{x}$ in cel B13 en D13. Vervolgens berekenen we in cel B15 en cel D15 de respectievelijke steekproefvarianties, waardoor we in cel B17 s_G^2 kunnen berekenen. Als we dan nog in cel B18 met s_G de populatiestandaarddeviatie σ hebben geschat, kunnen we de t-grootheid berekenen in cel B20:

$$t = \frac{(1\ 609 - 1\ 568) - (\mu_1 - \mu_2)}{232{,}33 \times \sqrt{\frac{1}{10} + \frac{1}{10}}}$$

$$= \frac{41}{232{,}33 \times 0{,}45}$$

$$= \frac{41}{104{,}55}$$

$$= +0{,}39$$

Met de Excelfunctie T. INV. 2T (T. INV. 2T) berekenen we de t-kritieke waarden voor de rechter- en linkerstaart van een t-verdeling met v = 20 – 2 en een α gelijk aan 5 %. In cel B22 vinden we +2,10 als kritieke waarde voor de rechterstaart, zodat -2,10 de kritieke waarde is in de linkerstaart.

Aangezien de berekende t-waarde kleiner is dan 2,10 weerhouden we H_0.

B22 | =T.INV.2T(0,05;18)

	A	B	C	D	E
1	Wedden onderneming A			Wedden onderneming B	
2	1350			1250	
3	1700			1750	
4	1850			1600	
5	2000			2100	
6	1800			1520	
7	1400			1500	
8	1475			1375	
9	1625			1725	
10	1500			1400	
11	1390			1460	
12					
13	$\bar{X}_A$	1609	$\bar{X}_B$	1568	
14					
15	s^2	49393,33	s^2	58556,67	
16					
17	s_g^2	53975			
18	s_g	232,33			
19					
20	t	0,39			
21					
22	t_k	2,10			
23					

Figuur 76 Testen van het verschil van $\bar{x}_1 - \bar{x}_2$ bij identieke populatievarianties

3.2 De ongekende populatievarianties zijn ongelijk

In dit geval gebruiken we de steekproefstandaarddeviaties s_1 en s_2 als schatters van σ_{x_1} en σ_{x_2}. De formule voor t wordt dan:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{s_{x_1}^2}{n_1} + \dfrac{s_{x_2}^2}{n_2}}}$$

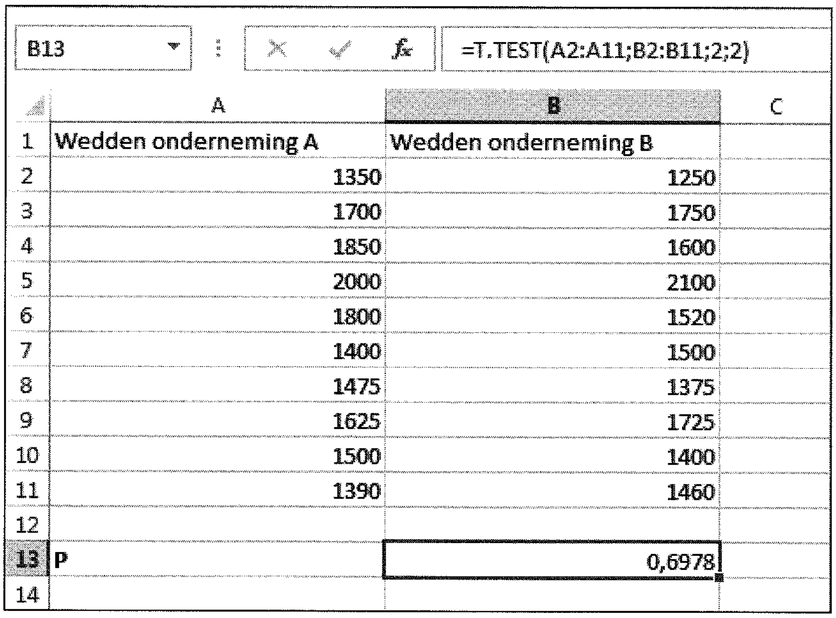
3.3 T. TEST (T. TEST)

Met de Excelfunctie T. TEST kunnen we het rekenwerk voor het uitvoeren van de t-test, die het verschil tussen twee steekproefgemiddelden test, achterwege laten.

In Figuur 77 hernemen we de steekproefgegevens van de wedden in de ondernemingen A en B en gebruiken nu de T. TEST:

> in kolom A hebben we het gegevensbereik A2:A11 met de weddegegevens van onderneming A en in kolom B het gegevensbereik B2:B11 met de weddegegevens van onderneming B;
> we kiezen cel B13 om het resultaat van de T. TEST te tonen;
> we klikken op de f_x-knop en selecteren in het venster 'Functie invoegen' de functie 'T. TEST';
> in het volgende venster 'Functieargumenten' voeren we in het vak 'Matrix 1' het gegevensbereik A2:A11 in;
> in het vak 'Matrix 2' voeren we de tweede gegevensverzameling B2:B11 in;
> in het vak 'Zijden' moeten we aanduiden of de toets eenzijdig (= 1) of tweezijdig (= 2) is. In ons geval is het een tweezijdige toets;
> in het vak 'Type-getal' duiden we aan welk type van t-test we willen uitvoeren. Er zijn drie mogelijkheden:
- een t-test voor gekoppelde steekproeven (= 1);
- een t-test voor twee steekproeven met veronderstelling van gelijke varianties (= 2);
- een t-test voor twee steekproeven met veronderstelling van ongelijke varianties (= 3).

We voeren code 2 in, sluiten het venster en lezen in cel B13 de kans af van t voor de steekproefgegevens. Omdat deze tweezijdige kans 0,6978 is en dus groter is dan 0,05; aanvaarden we H_0.



B13 | =T.TEST(A2:A11;B2:B11;2;2)

	A	B	C
1	Wedden onderneming A	Wedden onderneming B	
2	1350	1250	
3	1700	1750	
4	1850	1600	
5	2000	2100	
6	1800	1520	
7	1400	1500	
8	1475	1375	
9	1625	1725	
10	1500	1400	
11	1390	1460	
12			
13	P	0,6978	
14			

Figuur 77 De T. TEST

3.4 De T-toets als gegevensanalysefunctie

In Figuur 78 hernemen we de weddegegevens van de ondernemingen A en B en gebruiken we de toepassing T-toets: twee steekproeven met ongelijke varianties[2] (t-test: Two sample assuming unequal variances).

Deze functie wordt als volgt gebruikt:

> in kolom A voer je de steekproefgegevens in van onderneming A en in kolom B de steekproefgegevens van onderneming B;
> in het lint selecteer je het tabblad 'Gegevens'. Dan klik je op de knop 'Gegevensanalyse' om het dialoogvenster 'Gegevensanalyse' te openen;
> in het dialoogvenster 'Gegevensanalyse' selecteer je 'T-toets: twee steekproeven met ongelijke varianties' en klik je op 'OK' om het dialoogvenster van deze functie te openen;
> in het vak 'Variabelenbereik 1' voer je het bereik A2:A11 in, en in het vak 'Variabelenbereik 2' zet je het bereik B2:B11;
> in het vak 'Schatting van verschil tussen gemiddelden' voer je H_0 in en dat is $\mu_1 - \mu_2$ of m.a.w. 0;
> de optie 'Labels' wordt niet aangeduid aangezien we de kolomtitels niet hebben opgenomen in onze celbereiken;
> voor 'Alfa' vullen we 0,05 in;
> als Uitvoeroptie selecteren we 'Uitvoerbereik', voeren cel A15 in en de resultaten verschijnen in Figuur 78.

In cel B21 vinden we de H_0-hypothese en in cel B22 het aantal vrijheidsgraden. Cel B23 geeft de berekende t-waarde. Verder vinden we in cel B24 het oppervlak dat door de positieve t-waarde gelijk aan 0,39 wordt afgesneden in de bovenste staart van de t-verdeling. Cel B25 geeft ons de t-kritieke waarde voor een eenzijdige toets als α gelijk is aan 5 %. Cel B26 verdubbelt het oppervlak van cel B24, zijnde gelijk aan het oppervlak dat wordt afgesneden door de t-waarden +0,39 en -0,39 in beide staarten van de verdeling.
In cel B27 ten slotte lezen we de kritieke t-waarde af bij een tweezijdige toets, m.a.w. t = 2,10 snijdt 2,5 % af in de bovenste staart en t = -2,10 snijdt 2,5 % af in de onderste staart. Opnieuw kunnen we besluiten dat we H_0 aanvaarden, omdat de berekende t-waarde 0,39 kleiner is dan de kritieke t-waarde die gelijk is aan 2,10.

2 Bedoeld wordt dat de populatievarianties ongelijk zijn.

	A	B	C	D
1	Wedden onderneming A	Wedden onderneming B		
2	1350	1250		
3	1700	1750		
4	1850	1600		
5	2000	2100		
6	1800	1520		
7	1400	1500		
8	1475	1375		
9	1625	1725		
10	1500	1400		
11	1390	1460		
12				
13				
14				
15	T-toets: twee steekproeven met ongelijke varianties			
16				
17		*Variabele 1*	*Variabele 2*	
18	Gemiddelde	1609	1568	
19	Variantie	49393,33333	58556,66667	
20	Waarnemingen	10	10	
21	Schatting van verschil tussen gemiddelden	0		
22	Vrijheidsgraden	18		
23	T- statistische gegevens	0,39461404		
24	P(T<=t) eenzijdig	0,348882167		
25	Kritiek gebied van T-toets: eenzijdig	1,734063062		
26	P(T<=t) tweezijdig	0,697764334		
27	Kritiek gebied van T-toets: tweezijdig	2,100923666		
28				

Figuur 78 De T-toets: twee steekproeven met ongelijke varianties

4 Testen van gekoppelde steekproeven

We spreken van gekoppelde steekproeven als de gegevens betrekking hebben op eenzelfde steekproefelement.

Acht studenten nemen bijvoorbeeld deel aan een bijscholingsprogramma voor het vak informatica. Hun uitslagen op 10 punten op een informaticatoets vóór en na het bijscholingsprogramma en de verschillen tussen de behaalde punten tonen we in Figuur 79. De berekende verschillen D beschouwen we nu als een nieuwe steekproef waarmee we een eenzijdige hypothesetoets zullen uitvoeren met behulp van de t-verdeling. Daartoe berekenen we eerst $\bar{D}$, het gemiddelde van de verschillen, in cel B11 en vervolgens s_D in cel B12. Deze grootheden worden berekend met de Excelfuncties GEMIDDELDE en STDEV. S.

De hypothesen die we zullen testen, zijn:

H_0: $\mu_D \leq 0$

H_1: $\mu_D > 0$

$\alpha = 0{,}05$

De formule voor t is:

$$t = \frac{\bar{D} - \mu_D}{s_{\bar{D}}}$$

waarbij $s_{\bar{D}} = \dfrac{s_D}{\sqrt{n}}$

Toegepast op ons voorbeeld wordt dat:

$$t = \frac{0{,}625 - 0}{\frac{1{,}30}{\sqrt{8}}} = \frac{0{,}625}{0{,}46} = 1{,}36$$

Voor de functie T. INV (T. INV) geven we op dat v = 8 – 1 en dat de probabiliteit gelijk is aan 0,95; omdat we in dit voorbeeld te maken hebben met een eenzijdige test. Zo vinden we een kritieke t-waarde gelijk aan 1,89. Omdat t kleiner is dan de kritieke waarde van t, aanvaarden we H_0 en besluiten dat het bijscholingsprogramma geen verbetering van de uitslagen op de informaticatoets opleverde.

B14 =T.INV(0,95;7)

	A	B	C	D	E
1	**Student**	**Uitslag na bijscholing**	**Uitslag vóór bijscholing**	**Verschil D**	
2	1	5	4	1	
3	2	5	5	0	
4	3	8	6	2	
5	4	7	8	-1	
6	5	9	7	2	
7	6	6	6	0	
8	7	5	3	2	
9	8	8	9	-1	
10					
11	$\bar{D}$	**0,625**			
12	s_D	**1,30**			
13	t	**1,36**			
14	t_k	**1,89**			
15					

Figuur 79 Hypothesetoets voor gekoppelde steekproeven

4.1 T. TEST voor gekoppelde steekproeven

In Figuur 80 gebruiken we opnieuw de gegevens over de informaticatoets van vóór en na de bijscholing. Met die gegevens tonen we hoe de T. TEST gebruikt kan worden voor gekoppelde steekproeven:

> in kolom B voeren we de uitslagen in na bijscholing en in kolom C die van vóór de bijscholing;
> we selecteren cel B11;
> we klikken op de f_x-knop en openen het venster 'Functie invoegen';
> we selecteren de 'T. TEST' en openen het venster 'Functieargumenten';
> in het vak 'Matrix 1' voeren we het gegevensbereik B2:B9 in en in het vak 'Matrix 2' het bereik C2:C9;
> in het vak 'Zijden' voeren we het getal 1 in om aan te geven dat het een eenzijdige toets is;
> in het vak 'Type-getal' voeren we het getal 1 in om aan te geven dat het een T. TEST is voor gekoppelde steekproeven;
> we sluiten het venster 'Functieargumenten' en lezen in de geselecteerde cel B11 de kans van de t-waarde voor onze gegevens.

Omdat de kans dat t groter is dan de berekende t-waarde meer dan 5 % is, met name 11 %, aanvaarden we H_0 en besluiten we opnieuw dat de bijscholing geen beduidend effect heeft.

B11 | =T.TEST(B2:B9;C2:C9;1;1)

	A	B	C	D	E
1	Student	Uitslag na bijscholing	Uitslag vóór bijscholing	Verschil	
2	1	5	4	1	
3	2	5	5	0	
4	3	8	6	2	
5	4	7	8	-1	
6	5	9	7	2	
7	6	6	6	0	
8	7	5	3	2	
9	8	8	9	-1	
10					
11	Kans	0,1084			
12					

Figuur 80 T. TEST voor gekoppelde steekproeven

4.2 T-toets: twee gepaarde steekproeven voor gemiddelden (T-test: Paired two sample for means)

Nu illustreren we het gebruik van de T-toets: twee gepaarde steekproeven voor gemiddelden, een gegevensanalysetechniek die we toepassen op de punten behaald op de informaticatoets voor en na de bijscholing. Figuur 81 laat in het werkblad de gegevens ter zake zien. De T-toets verloopt als volgt:

> in kolom B hebben we de uitslagen na de bijscholing ingevoerd en in kolom C de uitslagen voor de bijscholing;
> in het lint selecteer je het tabblad 'Gegevens', klik je op de knop 'Gegevensanalyse' en het dialoogvenster 'Gegevensanalyse' wordt geopend;
> dan selecteer je de T-toets: twee gepaarde steekproeven voor gemiddelden en open je het betreffende venster;
> in het invoervak 'Variabelenbereik 1' voer je het bereik B2:B9 in;
> in het invoervak 'Variabelenbereik 2' zet je het bereik C2:C9;
> in het vak 'Schatting verschil tussen gemiddelden' voer je H_0 in. We stellen dus dat het verschil tussen μ_1 en μ_2 gelijk is aan nul;
> als de kolomtitels deel uitmaken van onze gegevensbereiken, tik je de optie 'Labels' aan, wat wij in dit geval echter niet doen;
> in het vak 'Alfa' is als foutenrisico 0,05 opgegeven;
> van de Uitvoeropties kies je de optie 'Uitvoerbereik', voer je cel A15 in, klik je op 'OK' en we zien de resultaten van onze toepassing.

In cel B19 verschijnt een nieuwe statistische grootheid: de *Pearson-correlatiecoëfficiënt*. Deze coëfficiënt duidt de samenhang aan tussen de gegevens uit de eerste en de tweede steekproef en ligt tussen -1 en +1. De coëfficiënt is gelijk aan 0,76, wat wijst op een vrij sterke positieve samenhang tussen de gegevens van beide steekproeven.
Cel B20 toont de H_0-hypothese en cel B21 het aantal vrijheidsgraden. In cel B22 lezen we de berekende t-waarde af en daarmee samenhangend in cel B23 het oppervlak dat de t-waarde afsnijdt in de bovenste staart van de t-verdeling.
Cel B24 geeft ons de kritieke t-waarde voor een eenzijdige toets die een oppervlak van 5 % afsnijdt in de bovenste staart.
Cel B25 geeft ons het oppervlak dat de t-waarden +1,36 en -1,36 afsnijden in respectievelijk de bovenste en onderste staart van de t-verdeling. Cel B26 toont de kritieke t-waarde voor een tweezijdige toets. Dat betekent dat de kritieke waarden +2,36 en -2,36 elk 2,5 % afsnijden in de bovenste en onderste staart van de t-verdeling. Het is duidelijk dat ook met deze toets H_0 niet wordt verworpen, aangezien 1,36 kleiner is dan 1,89.

	A	B	C	D
1	Student	Uitslag na bijscholing	Uitslag vóór bijscholing	
2	1	5	4	
3	2	5	5	
4	3	8	6	
5	4	7	8	
6	5	9	7	
7	6	6	6	
8	7	5	3	
9	8	8	9	
10				
11				
12				
13	T-toets: twee gepaarde steekproeven voor gemiddelden			
14				
15		*Variabele 1*	*Variabele 2*	
16	Gemiddelde	6,625	6	
17	Variantie	2,553571429	4	
18	Waarnemingen	8	8	
19	Pearson-correlatie	0,759883266		
20	Schatting van verschil tussen gemiddelden	0		
21	Vrijheidsgraden	7		
22	T- statistische gegevens	1,357241785		
23	P(T<=t) eenzijdig	0,108418773		
24	Kritiek gebied van T-toets: eenzijdig	1,894577508		
25	P(T<=t) tweezijdig	0,216837546		
26	Kritiek gebied van T-toets: tweezijdig	2,36462256		
27				

Figuur 81 T-toets: twee gepaarde steekproeven voor gemiddelden

5 Testen van varianties

Aan de hand van het volgende voorbeeld tonen we hoe we varianties kunnen testen. We nemen aan dat in een industriële onderneming kleine metalen ringen worden geproduceerd die slechts een kleine variabiliteit mogen hebben inzake diameter. De productieafdeling bestudeert twee verschillende machines die zulke ringen maken en wil weten welke van beide de kleinste variabiliteit genereert t.o.v. de gemiddelde diameter van de ringen in beide populaties. Met machine A worden 10 ringen geproduceerd en s_A^2 is gelijk aan 0,04678 mm^2. Met machine B worden 12 ringen geproduceerd en s_B^2 is 0,03295 mm^2. De formulering van de hypothesen voor deze tweezijdige test is:

$$H_0: \sigma_A^2 = \sigma_B^2$$

$$H_1: \sigma_A^2 \neq \sigma_B^2$$

$$\alpha = 0{,}05$$

Om de varianties te testen, gebruiken we een nieuwe toetsingsgrootheid, de F-ratio:

$$F = \frac{s_A^2}{s_B^2}$$

De F-ratio heeft een F-verdeling met $n_A - 1$ vrijheidsgraden in de teller en $n_B - 1$ vrijheidsgraden in de noemer, omdat bij de berekening van F twee variantieschatters s_A^2 en s_B^2 nodig zijn.
We berekenen F en vinden:

$$F = \frac{0{,}04678}{0{,}03295} = 1{,}42$$

We vergelijken deze F-waarde met de kritieke F-waarde voor 9 en 11 vrijheidsgraden en een kans = 0,975. Met behulp van de Excelfunctie F. INV vinden we een kritieke waarde van 3,59. We besluiten dat H_0 niet wordt verworpen of dus dat het verschil tussen beide steekproefvarianties onbeduidend is.

5.1 F. TEST (F. TEST)

Deze Excelfunctie berekent de tweezijdige kans van de berekende F-ratio als H_0 waar is. Maar deze F-waarde – in ons voorbeeld 1,42 – wordt met deze functie niet getoond.

In Figuur 82 tonen we de diameters van de ringen geproduceerd door de twee machines in de steekproeven, alsook de steekproefgemiddelden en de steekproefvarianties.

We voeren de F. TEST nu als volgt uit:

> voer de gegevens voor machine A in kolom B in en de gegevens voor machine B in kolom D;
> selecteer cel D21 voor het resultaat;
> klik op de f_x-knop en open het venster 'Functie invoegen', selecteer de 'F. TEST' en open het venster 'Functieargumenten';
> in het vak 'Matrix 1' voer je het gegevensbereik B3:B12 in, en in het vak 'Matrix 2' het gegevensbereik D3:D14;
> je sluit het venster door op 'OK' te klikken en we lezen in cel D21 het resultaat.

Omdat 0,58 groter is dan 0,05, verwerpen we H_0 ook dit keer niet.

D21 =F.TEST(B3:B12;D3:D14)

	A	B	C	D	E	F
1						
2		Machine A		Machine B		
3		5		4,9		
4		5,2		5,1		
5		5,1		4,7		
6		4,8		4,9		
7		5,1		5		
8		4,8		4,8		
9		5,2		5,1		
10		4,6		4,8		
11		4,7		4,6		
12		4,8		4,9		
13				5,1		
14				5,2		
15						
16						
17	$\overline{X}A$	4,93	$\overline{X}B$	4,93		
18						
19	s^2_A	0,04678	s^2_B	0,03295		
20						
21			P	0,58		
22						

Figuur 82 De F. TEST voor het testen van varianties

5.2 F. VERD (F. DIST)

Met de functie F. VERD kunnen we uitmaken of een berekende F-waarde zich al of niet in het verwerpingsgebied bevindt. Zij berekent de kans dat de F-waarde kleiner of gelijk is aan een opgegeven waarde van F. Als die kans groter is dan 1 verminderd met het foutenrisico in kwestie, kan H_0 niet worden weerhouden.

We passen de functie toe op het voorbeeld van de diameters van geproduceerde ringen. De F-ratio is 1,42 met 9 vrijheidsgraden in de teller en 11 vrijheidsgraden in de noemer. Figuur 83 toont opnieuw het werkblad met de steekproefgegevens voor machine A en machine B.
De functie voeren we als volgt uit:

> we selecteren cel B16 voor het resultaat;
> we klikken op de f_x-knop en openen het venster 'Functie invoegen', selecteren de 'F. VERD' en openen het venster 'Functieargumenten';

> in het X-vak voeren we de berekende F-waarde in, dat is 1,42;
> in het vak 'Vrijheidsgraden 1' zetten we het aantal vrijheidsgraden van de teller, zijnde 9;
> in het vak 'Vrijheidsgraden 2' zetten we het aantal vrijheidsgraden van de noemer, zijnde 11;
> in het laatste vak 'Cumulatief' typen we 'WAAR' in;
> we sluiten het venster 'Functieargumenten' en het resultaat verschijnt in cel B16. De F-waarde van 1,42 of kleiner heeft een kans van 0,71 als H_0 waar is. Omdat deze kans kleiner is dan 1 verminderd met het tweezijdige foutenrisico van α = 0,025, aanvaarden we H_0. Of, met andere woorden, de diameters van de ringen geproduceerd door machine A variëren niet meer of minder dan die geproduceerd door machine B.

B16 =F.VERD(1,42;9;11;WAAR)

	A	B	C	D	E	F
1						
2	**Machine A**	**Machine B**				
3	5	4,9				
4	5,2	5,1				
5	5,1	4,7				
6	4,8	4,9				
7	5,1	5				
8	4,8	4,8				
9	5,2	5,1				
10	4,6	4,8				
11	4,7	4,6				
12	4,8	4,9				
13		5,1				
14		5,2				
15						
16	**P**	**0,7127**				
17						

Figuur 83 De F.VERD-functie

5.3 F. INV. RECHTS (F. INV. RT)

De werkbladfunctie F. INV. RECHTS berekent de F-waarde in de F-verdeling die een vooraf opgegeven oppervlak afsnijdt in de bovenste staart. We gebruiken dus deze functie om de kritieke F-waarde te vinden.

Toegepast op het voorbeeld rond de variabiliteit van de diameter van de ringen geproduceerd door machine A en machine B, gebruiken we de functie om de kritieke waarde te vinden voor de tweezijdige test:

- in Figuur 84 selecteren we in het werkblad cel B16 om de kritieke F-waarde in te tonen;
- we klikken op de f_x-knop, openen het venster 'Functie invoegen', selecteren 'F. INV. RECHTS' en openen het venster 'Functieargumenten';
- in het Kans-vak voeren we het oppervlak in dat de kritieke F-waarde zal afsnijden in de bovenste staart. Omdat het een tweezijdige toets is en $\alpha = 0{,}05$, moeten we 0,025 intikken;
- in het vak 'Vrijheidsgraden 1' voeren we de vrijheidsgraden in van de teller, zijnde 9;
- in het vak 'Vrijheidsgraden 2' voeren we de vrijheidsgraden in van de noemer, zijnde 11;
- we sluiten het venster en in cel B16 verschijnt de kritieke F-waarde van 3,59.

Omdat de kritieke F-waarde groter is dan de berekende F-waarde van 1,42, aanvaarden we H_0.

B16 | =F.INV.RECHTS(0,025;9;11)

	A	B	C	D	E	F
1						
2	**Machine A**	**Machine B**				
3	5	4,9				
4	5,2	5,1				
5	5,1	4,7				
6	4,8	4,9				
7	5,1	5				
8	4,8	4,8				
9	5,2	5,1				
10	4,6	4,8				
11	4,7	4,6				
12	4,8	4,9				
13		5,1				
14		5,2				
15						
16	F_k	3,59				
17						

Figuur 84 De F. INV. RECHTS-functie

5.4 F-toets voor twee steekproeven (F-Test, two sample for variances)

Met deze functie beschikken we over een gegevensanalysefunctie om een F-test op twee steekproefvarianties uit te voeren.
Figuur 85 toont opnieuw de steekproefgegevens die nodig zijn om de toepassing te kunnen gebruiken:

> in kolom A voeren we de diametergegevens in van de ringen geproduceerd door machine A en in kolom B de diametergegevens van de ringen geproduceerd door machine B;
> in het lint selecteren we het tabblad 'Gegevens' en klikken nadien op de knop 'Gegevensanalyse'. In het venster dat verschijnt, kiezen we de 'F-toets voor twee steekproeven' en klikken op 'OK';
> in het dialoogvenster F-toets voor twee steekproeven voeren we in het vak 'Variabelenbereik 1' het gegevensbereik A3:A12 in, en in het vak 'Variabelenbereik 2' het gegevensbereik B3:B14 in;
> de optie 'Labels' wordt niet aangeduid omdat de kolomtitels geen deel uitmaken van de gegevensbereiken;
> in het vak 'Alfa' voeren we een foutenrisico van 0,025 in, omdat het gaat over een tweezijdige toets;
> als outputoptie kiezen we 'Uitvoerbereik' en duiden aan waar we de resultatentabel willen;
> we klikken op 'OK' en zien de resultaten verschijnen in ons werkblad.

In cel B23 lezen we de berekende F-waarde af en in cel B24 vinden we het oppervlak dat de berekende F-waarde afsnijdt in de bovenste staart van de F-verdeling. Cel B25 toont de kritieke waarde van F voor een eenzijdige toets die 2,5 % afsnijdt in de bovenste staart. Omdat de F-waarde 1,42 kleiner is dan de kritieke F-waarde 3,59, kunnen we opnieuw besluiten dat we de nulhypothese niet verwerpen.

	A	B	C	D
1				
2	Machine A	Machine B		
3	5	4,9		
4	5,2	5,1		
5	5,1	4,7		
6	4,8	4,9		
7	5,1	5		
8	4,8	4,8		
9	5,2	5,1		
10	4,6	4,8		
11	4,7	4,6		
12	4,8	4,9		
13		5,1		
14		5,2		
15				
16	F-toets: twee steekproeven voor varianties			
17				
18		*Variabele 1*	*Variabele 2*	
19	Gemiddelde	4,93	4,925	
20	Variantie	0,046777778	0,032954545	
21	Waarnemingen	10	12	
22	Vrijheidsgraden	9	11	
23	F	1,419463602		
24	P(F<=f) eenzijdig	0,287513495		
25	Kritisch gebied van F-toets: eenzijdig	3,587899755		
26				

Figuur 85 F-toets voor twee steekproeven

hoofdstuk 16

Variantieanalyse

1 Unifactoriële variantieanalyse

We behandelen nu een methode voor het testen van hypothesen m.b.t. drie of meer steekproeven.

Veronderstel dat we in een verffabriek drie afvullijnen moeten vergelijken. We willen namelijk weten of de verschillen in het aantal afgevulde potten verf per uur door de drie afvullijnen al of niet toevallig zijn.

In Figuur 86 vinden we voor 5 verschillende afvuluren de resultaten opgeleverd door de 3 afvullijnen en ook een aantal statistische grootheden. Om uit te maken of de resultaten al of niet verschillend zijn, formuleren we de volgende hypothesen:

H_0: $\mu_1 = \mu_2 = \mu_3$

H_1: H_0 is niet geldig

$\alpha = 0{,}05$

	A	B	C	D	E
1		Afvullijn 1 (in 00)	Afvullijn 2 (in 00)	Afvullijn 3 (in 00)	
2		37	33	27	
3		36	34	30	
4		34	33	31	
5		35	35	29	
6		38	32	30	
7					
8					
9	Gemiddelde	36,00	33,40	29,40	
10					
11	Variantie	2,50	1,30	2,30	
12					
13	Standaarddeviatie	1,58	1,14	1,52	
14					
15	Gemiddelde van de steekproefgemiddelden	32,93			
16					

Figuur 86 Steekproefgegevens van drie afvullijnen

Om deze hypothesen te testen, gebruiken we de variantieanalysetechniek. Deze techniek is gebaseerd op de berekening van twee soorten varianties. Ten eerste is er de berekening van de variantie tussen de steekproefgemiddelden. De formule voor deze variantie is:

$$s_{\bar{x}}^2 = \frac{\Sigma n_i \times (\bar{X}_i - \bar{\bar{X}})^2}{c - 1}$$

c = aantal lijnen

c - 1 } = vrijheidsgraden

aantal observaties

VARIANTIEANALYSE

waarbij n_i staat voor het aantal steekproefelementen in de i-de steekproef, c staat voor het aantal steekproeven en $\bar{\bar{x}}$ het gemiddelde voorstelt van de steekproefgemiddelden $\bar{x}_i$. Als we deze variantie toepassen op het voorbeeld, krijgen we:

$$s_{\bar{x}}^2 = \frac{5 \times (36{,}00 - 32{,}93)^2 + 5 \times (33{,}40 - 32{,}93)^2 + 5 \times (29{,}40 - 32{,}93)^2}{3 - 1}$$

$$= \frac{47{,}12 + 1{,}10 + 62{,}30}{2}$$

$$= 55{,}26$$

Een tweede berekening betreft de variantie in de steekproeven. De formule voor deze variantie is:

$$s_G^2 = \frac{(n_1 - 1) \times s_1^2 + (n_2 - 1) \times s_2^2 + (n_3 - 1) \times s_3^2}{(n_1 - 1) + (n_2 - 1) + (n_3 - 1)}$$

Toegepast op ons voorbeeld is dat:

$$s_G^2 = \frac{(5 - 1) \times 2{,}50 + (5 - 1) \times 1{,}30 + (5 - 1) \times 2{,}30}{4 + 4 + 4}$$

$$= \frac{10 + 5{,}20 + 9{,}20}{12}$$

$$= 2{,}03$$

De vraag die nu rijst, is of de verschillen tussen de steekproefgemiddelden al of niet toevallig zijn. Voor het antwoord op deze vraag vergelijken we de twee varianties en geldt de redenering dat als de variantie tussen de steekproeven beduidend groter is dan de variantie in de steekproeven, de verschillen in de steekproefgemiddelden dan het gevolg zijn van een verschil tussen de 3 afvullijnen. We moeten dan H_0 verwerpen. In het omgekeerde geval, als beide varianties ongeveer even groot zijn of als de variantie tussen de steekproeven kleiner is dan de variantie in de steekproeven, aanvaarden we H_0 en besluiten dat de verschillen in steekproefgemiddelden toevallig zijn.
Statistisch gezien gebeurt deze vergelijking door de berekening van de F-ratio:

$$F = \frac{s_{\bar{x}}^2}{s_G^2}$$

Voor de berekende F-waarde wordt dan verder nagegaan welk oppervlak zij afsnijdt in de bovenste staart van de specifieke F-verdeling die bepaald wordt door het aantal vrijheidsgraden in teller en noemer van de F-ratio.

Voor ons voorbeeld is:

$$F = \frac{55,26}{2,03}$$

$$= 27,22$$

Voor 2 vrijheidsgraden in de teller en 12 in de noemer is het afgesneden oppervlak door de berekende F-waarde in de bovenste staart van deze verdeling kleiner dan 0,001 (zie bijlage Tabellen). We kunnen H_0 dus verwerpen, omdat F in het verwerpingsgebied α ligt.

2 Gegevensanalysefunctie: unifactoriële variantieanalyse (Anova: Single Factor)

In Figuur 87 hernemen we de steekproefgegevens uit het vorige voorbeeld en gebruiken nu voor onze variantieanalyse de gegevensanalysefunctie 'Unifactoriële variantieanalyse'.

Deze functie gebruiken we als volgt:

> we voeren de steekproefgegevens in voor de drie afvullijnen;
> in het lint selecteren we het tabblad 'Gegevens', klikken we op de knop 'Gegevensanalyse' en het betreffende dialoogvenster verschijnt;
> in het dialoogvenster 'Gegevensanalyse' kiezen we de functie 'Unifactoriële variantieanalyse';
> in het dialoogvenster 'Unifactoriële variantieanalyse' voeren we in het vak 'Invoerbereik' het gegevensbereik B2:D6 in;
> het gegevensbereik bevat geen kolomtitels, dus moeten we 'Labels' in de eerste rij niet selecteren;
> in het vak 'Alfa' is het foutenrisico 0,05 automatisch opgegeven;
> als Uitvoeroptie kiezen we 'Uitvoerbereik' en geven op waar de resultaten moeten verschijnen;
> we klikken op 'OK' en de resultaten verschijnen in het werkblad.

In de eerste tabel *Samenvatting* vinden we statistische grootheden van de drie steekproeven. In de tabel *Variantieanalyse* wordt in de kolom *Gemiddelde kwadraten* de variantie tussen de steekproefgemiddelden getoond en de variantie in de steekproeven. Ook lezen we de berekening van F en de kans dat F groter of gelijk is aan 27,18 of m.a.w. het oppervlak dat F afsnijdt in de bovenste staart van de F-verdeling. Ten slotte vinden we er ook nog de kritieke F-waarde. Omdat F groter is dan 3,89, verwerpen we H_0.

	A	B	C	D	E	F	G	H
1		Afvullijn 1 (in 00)	Afvullijn 2 (in 00)	Afvullijn 3 (in 00)				
2		37	33	27				
3		36	34	30				
4		34	33	31				
5		35	35	29				
6		38	32	30				
7								
8	Unifactoriële variantie-analyse							
9								
10	SAMENVATTING							
11	*Groepen*	*Aantal*	*Som*	*Gemiddelde*	*Variantie*			
12	Kolom 1	5	180	36	2,5			
13	Kolom 2	5	167	33,4	1,3			
14	Kolom 3	5	147	29,4	2,3			
15								
16								
17	Variantie-analyse							
18	*Bron van variatie*	*Kwadratensom*	*Vrijheidsgraden*	*Gemiddelde kwadraten*	*F*	*P-waarde*	*Kritische gebied van F-toets*	
19	Tussen groepen	110,53	2	55,27	27,18	3,5E-05	3,89	
20	Binnen groepen	24,4	12	2,03				
21								
22	Totaal	134,93	14					
23								

Figuur 87 Resultaten van de gegevensanalysefunctie Unifactoriële variantieanalyse

3 Tweefactorvariantieanalyse

In het voorbeeld van de afvullijnen hebben we getest of er tussen de drie afvullijnen een verschil bestaat. Maar in vele gevallen moeten we meer dan één factor testen. Dat kunnen we doen door het uitvoeren van een tweefactorvariantieanalyse.

We hernemen het voorbeeld van de afvullijnen en classificeren de gegevens nu naar twee factoren, te weten naar afvullijn en naar arbeidsploeg. In Figuur 88 zien we het resultaat van deze nieuwe ordening samen met de berekening van de arbeidsploeggemiddelden, de afvullijngemiddelden en het gemiddelde van de steekproefgemiddelden.

	A	B	C	D	E	F
1						
2	Arbeidsploeg		Afvullijn (in 00)		Gemiddelde	
3		1	2	3		
4	1	37	33	27	32,33	
5	2	36	34	30	33,33	
6	3	34	33	31	32,67	
7	4	35	35	29	33,00	
8	5	38	32	30	33,33	
9	Gemiddelde	36,00	33,40	29,40	32,93	
10						

Figuur 88 Steekproefgegevens voor de drie afvullijnen geordend volgens arbeidsploeg

De formulering van de hypothesetesten is als volgt:

H_0: $\mu_{\text{afvullijn } 1} = \mu_{\text{afvullijn } 2} = \mu_{\text{afvullijn } 3}$

H_1: H_0 verwerpen

en

H_0: $\mu_{\text{arbpl } 1} = \mu_{\text{arbpl } 2} = \mu_{\text{arbpl } 3} = \mu_{\text{arbpl } 4} = \mu_{\text{arbpl } 5}$

H_1: H_0 verwerpen

α is voor beide toetsen gelijk aan 0,05.

Om deze testen uit te voeren, berekenen we achtereenvolgens:

a de variantie tussen de afvullijnen (= k)

$$s^2_{\bar{x}_{i.}} = \frac{r \sum_{i=1}^{k} (\bar{x}_{i.} - \bar{\bar{x}})^2}{k - 1}$$

$$= \frac{5\,(36 - 32{,}93)^2 + 5\,(33{,}40 - 32{,}93)^2 + 5\,(29{,}40 - 32{,}93)^2}{2}$$

$$= \frac{47{,}125 + 1{,}105 + 62{,}305}{2}$$

$$= \frac{110{,}535}{2}$$

$$= 55{,}268$$

b de variantie tussen de arbeidsploegen (= r)

$$s^2_{\bar{x}_{.j}} = \frac{k \sum_{j=1}^{r} (\bar{x}_{.j} - \bar{\bar{x}})^2}{r - 1}$$

$$= \frac{3\,(32{,}33 - 32{,}93)^2 + 3\,(33{,}33 - 32{,}93)^2 + 3\,(32{,}67 - 32{,}93)^2 + 3\,(33{,}00 - 32{,}93)^2 + 3\,(33{,}33 - 32{,}93)^2}{4}$$

$$= \frac{1{,}080 + 0{,}480 + 0{,}203 + 0{,}015 + 0{,}480}{4}$$

$$= \frac{2{,}258}{4}$$

$$= 0{,}565$$

c de toevalsvariantie

Om deze variantie te berekenen, becijferen we eerst de totale variantie of de variantie in de steekproefgegevens:

$$s^2_{x_{ij}} = \frac{\Sigma\,(x_{ij} - \bar{\bar{x}})^2}{v}$$

$$= \frac{(37 - 32{,}93)^2 + (33 - 32{,}93)^2 + (27 - 32{,}93)^2 + \ldots + (30 - 32{,}93)^2}{14}$$

$$= \frac{16{,}565 + 0{,}005 + 35{,}165 + 9{,}425 + 1{,}145 + 8{,}585 + 1{,}145 + 0{,}005 + 3{,}725 + 4{,}285 + 4{,}285 + 15{,}445 + 25{,}705 + 0{,}865 + 8{,}585}{14}$$

$$= \frac{134{,}935}{14}$$

$$= 9{,}638$$

De toevalsvariantie is dan:

$$\text{Toevalsvariantie} = \frac{\Sigma\,(x_{ij} - \bar{\bar{x}})^2 - r\sum_{i=1}^{k}(\bar{x}_{i.} - \bar{\bar{x}})^2 - k\sum_{j=1}^{r}(\bar{x}_{.j} - \bar{\bar{x}})^2}{(n-1)-(k-1)-(r-1)}$$

$$= \frac{134{,}935 - 110{,}535 - 2{,}258}{14 - 2 - 4}$$

$$= \frac{22{,}142}{8}$$

$$= 2{,}768$$

Nu kunnen we testen of de verschillen tussen afvullijnen en arbeidsploegen al of niet beduidend zijn. De F-ratio om de verschillen tussen de afvullijnen te testen is:

$$F = \frac{s^2_{\bar{x}_{i.}}}{\text{toevalsvariantie}}$$

$$= \frac{55{,}268}{2{,}768}$$

$$= 19{,}967$$

Met 2 en 8 vrijheidsgraden en $\alpha = 0{,}05$ is de kritieke F-waarde gelijk aan 4,46 (zie bijlage Tabellen). Dat betekent dat we H_0 verwerpen voor de afvullijnen.
Vervolgens berekenen we de F-ratio om het verschil tussen de arbeidsploegen te evalueren:

$$F = \frac{s^2_{\bar{x}_{.j}}}{\text{toevalsvariantie}}$$

$$= \frac{0{,}565}{2{,}768}$$

$$= 0{,}204$$

Met 4 vrijheidsgraden in de teller en 8 in de noemer en een $\alpha = 0{,}05$, is de kritieke waarde van F hier gelijk aan 3,84 (zie bijlage Tabellen). Dat houdt in dat we H_0 niet verwerpen en dat er dus geen significant verschil is tussen de arbeidsploegen.

4 Gegevensanalysefunctie: multifactoriële variantieanalyse zonder herhaling (Anova: Two Factor without Replication)

De hierboven uitgevoerde analyse kunnen we ook doen met de Excelfunctie 'multifactoriële variantieanalyse zonder herhaling'. Met 'zonder herhaling' wordt bedoeld dat er van iedere combinatie afvullijn – arbeidsploeg slechts één observatie werd waargenomen.

Figuur 89 toont enerzijds de gegevens en anderzijds de resultaten van de multifactoriële analyse. De functie wordt in de volgende stappen uitgevoerd:

> je voert in het bereik A2:D8 de gegevens in;
> in het lint selecteer je het tabblad 'Gegevens', dan klik je op de knop 'Gegevensanalyse' en in het venster 'Gegevensanalyse' selecteer je de functie 'Multifactoriële variantieanalyse zonder herhaling'. Je klikt op 'OK' en het venster van de geselecteerde functie verschijnt;
> in het vak 'Invoerbereik' voer je het gegevensbereik B4:D8 in;
> 'Labels' gebruik je niet, omdat er geen kolomtitels in het gegevensbereik zijn opgenomen;
> in het vak 'Alfa' is het foutenrisico al ingevoerd;
> als Uitvoeroptie selecteer je 'Uitvoerbereik' en geef je in waar de resultaten moeten verschijnen;
> je klikt op 'OK' en we vinden de resultaten in het werkblad;

De resultaten bestaan uit twee tabellen. De tabel *Samenvatting* bevat een aantal gegevens over de rijen en kolommen. De tabel *Variantieanalyse* toont in de kolom *Gemiddelde kwadraten* achtereenvolgens de variantie tussen de arbeidsploegen, de variantie tussen de afvullijnen en de toevalsvariantie. Verder vinden we er ook de F-waarden voor het testen van de verschillen tussen de arbeidsploegen en de verschillen tussen de afvullijnen. Ten slotte vinden we er ook de P-waarden als H_0 waar is en de betreffende kritische F-waarden.
Op basis van de resultaten beslissen we H_0 te aanvaarden voor de arbeidsploegen en H_0 te verwerpen voor de afvullijnen.

	A	B	C	D	E	F	G	H
1								
2	Arbeidsploeg		Afvullijn (in 00)					
3		1	2	3				
4	1	37	33	27				
5	2	36	34	30				
6	3	34	33	31				
7	4	35	35	29				
8	5	38	32	30				
9								
10								
11	Multifactoriële analyse zonder herhaling							
12								
13	*SAMENVATTING*	*Aantal*	*Som*	*Gemiddelde*	*Variantie*			
14	Rij 1	3	97	32,33333333	25,3333333			
15	Rij 2	3	100	33,33333333	9,33333333			
16	Rij 3	3	98	32,66666667	2,33333333			
17	Rij 4	3	99	33	12			
18	Rij 5	3	100	33,33333333	17,3333333			
19								
20	Kolom 1	5	180	36	2,5			
21	Kolom 2	5	167	33,4	1,3			
22	Kolom 3	5	147	29,4	2,3			
23								
24								
25	Variantie-analyse							
26	*Bron van variatie*	*Kwadratensom*	*Vrijheidsgraden*	*Gemiddelde kwadraten*	*F*	*P-waarde*	*Kritische gebied van F-toets*	
27	Rijen	2,266666667	4	0,566666667	0,20481928	0,928647	3,837854479	
28	Kolommen	110,5333333	2	55,26666667	19,9759036	0,000775	4,458968306	
29	Fout	22,13333333	8	2,766666667				
30								
31	Totaal	134,9333333	14					
32	Totaal	135,6888889	19					
33								

Figuur 89 Resultaten van de gegevensanalysefunctie Multifactoriële analyse zonder herhaling

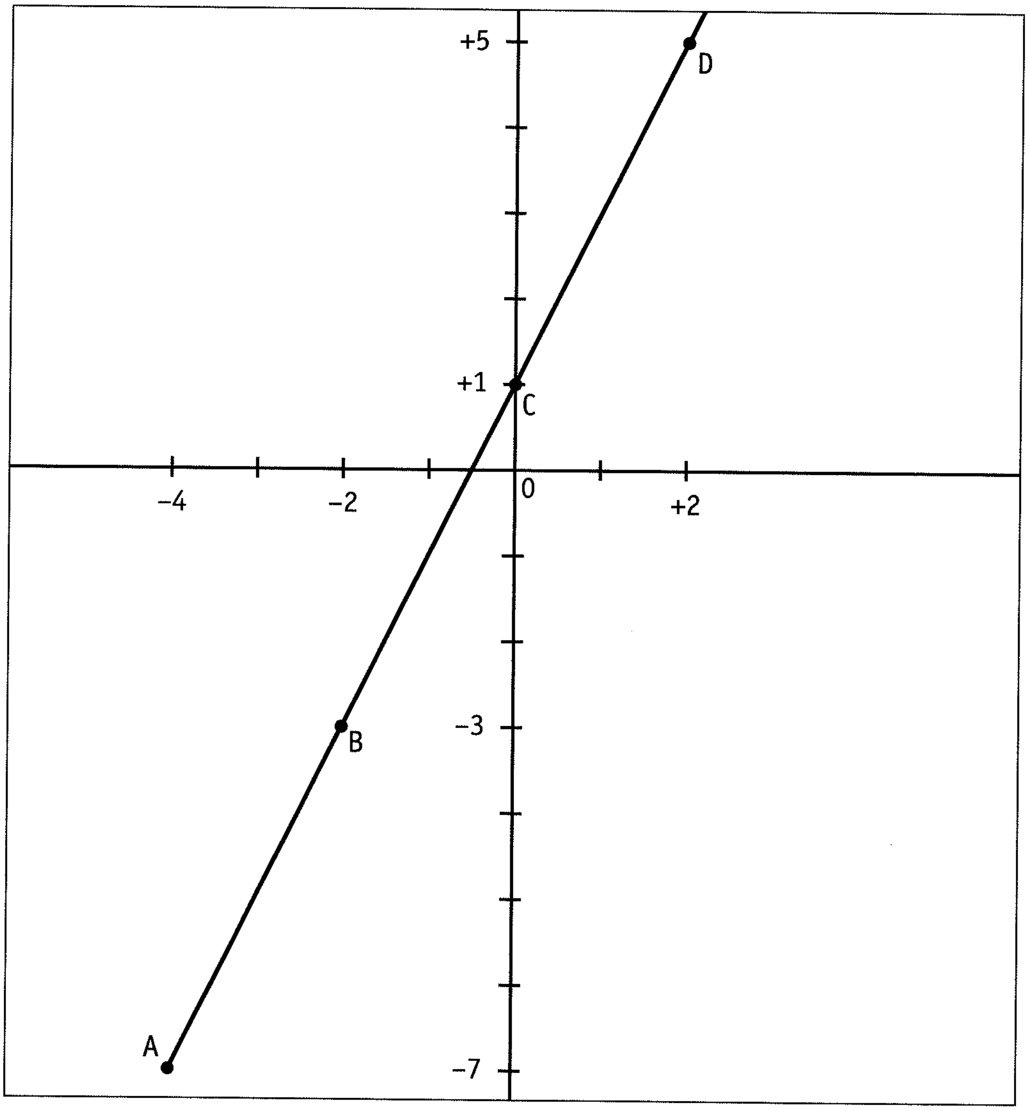
hoofdstuk 17

Regressie

1 De rechte lijn

Een rechte lijn is de grafische voorstelling van een relatie tussen een onafhankelijke variabele x en een afhankelijke variabele y. De vergelijking van een rechte lijn is bijvoorbeeld y = 2x + 1.
Deze vergelijking kunnen we tekenen door van een aantal (x, y) coördinaten de punten af te beelden in een coördinatenstelsel. Het resultaat zie je in Figuur 90:

x	y = 2x+1
-4	-7 → A (-4, -7)
-2	-3 → B (-2, -3)
0	+1 → C (0, +1)
+2	+5 → D (+2, +5)



Figuur 90 Grafiek van de vergelijking y = 2x + 1

De helling van de lijn wordt weergegeven door de richtingscoëfficiënt en geeft aan met hoeveel y verandert als x verandert met 1 eenheid. In ons voorbeeld is de richtingscoëfficiënt 2, dit betekent dat als x verandert met 1 eenheid, y zal veranderen met 2 eenheden.
Als x = 0, dan is y = 1 en 1 is de lengte van het lijnstuk dat de grafiek van de y-as afsnijdt.
Algemeen wordt de vergelijking van een rechte voorgesteld door: y = ax + b.

2 Enkelvoudige lineaire regressie

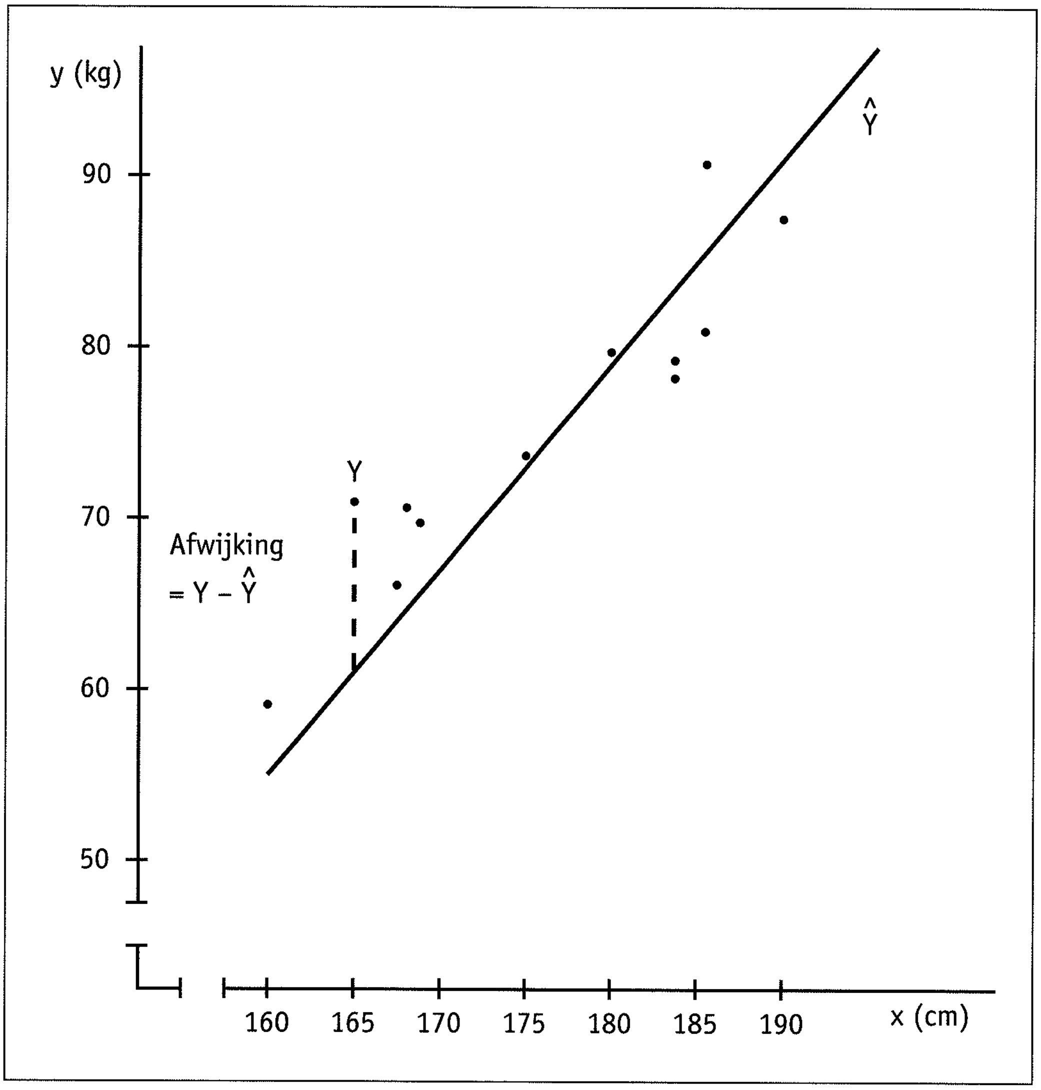
2.1 Regressievergelijking

Regressie is een techniek die in de statistiek gebruikt wordt om, in dit geval, een mogelijk verband tussen twee variabelen op te sporen.

Ter illustratie beschouwen we de lengten X en de gewichten Y van 12 studenten die lukraak geselecteerd werden op een hogeschool:

Student	Lente in cm: X	Gewicht in kg: Y
1	165	71
2	170	70
3	180	79
4	168	71
5	184	80
6	168	67
7	184	79
8	160	58
9	182	88
10	190	84
11	187	82
12	175	76

In de onderstaande Figuur 91 tekenen we de punten (x, y) in het rechthoekige assenkruis:



Figuur 91 Benaderende rechte

Nu trekken we door de puntenwolk in Figuur 91 een lijn die zo goed mogelijk de werkelijke punten benadert. Deze lijn noemen we de benaderende rechte. Deze rechte heeft als vergelijking:

$\hat{y} = ax + b$

waarbij:

$\hat{y}$ = geschatte waarde van y

Om bovenstaande vergelijking op te stellen, moeten we de regressiecoëfficiënten a en b berekenen. Het komt er nu op aan waarden voor a en b te vinden, zodat de verticale y-afwijkingen in Figuur 91 minimaal zijn.

Dat doen we via de volgende formules:

$$a = \frac{\Sigma\,(x - \bar{x})\,(y - \bar{y})}{\Sigma\,(x - \bar{x})^2}$$

$$b = \bar{y} - a\bar{x}$$

Dit passen we toe in het Excelwerkblad van Figuur 92.

	A	B	C	D	E	F	G	H
1								
2		X	Y	$X-\bar{X}$	$Y-\bar{Y}$	$(X-\bar{X})*(Y-\bar{Y})$	$(X-\bar{X})$^2	
3		165	71	-11,08	-4,42	48,95	122,84	
4		170	70	-6,08	-5,42	32,95	37,01	
5		180	79	3,92	3,58	14,03	15,34	
6		168	71	-8,08	-4,42	35,70	65,34	
7		184	80	7,92	4,58	36,28	62,67	
8		168	67	-8,08	-8,42	68,03	65,34	
9		184	79	7,92	3,58	28,37	62,67	
10		160	58	-16,08	-17,42	280,12	258,67	
11		182	88	5,92	12,58	74,45	35,01	
12		190	84	13,92	8,58	119,45	193,67	
13		187	82	10,92	6,58	71,87	119,17	
14		175	76	-1,08	0,58	-0,63	1,17	
15								
16	$\bar{X}$	176,08						
17	$\bar{Y}$	75,42						
18	$\Sigma\,(X-\bar{X})*(Y-\bar{Y})$	809,58						
19	$\Sigma\,(X-\bar{X})$^2	1038,92						
20	a	0,779						
21	b	-61,798						
22								

Figuur 92 Berekening van de regressiecoëfficiënten a en b

Zo vinden we de vergelijking van de lijn die zo dicht mogelijk bij de werkelijke waarden van y komt:

$$\hat{y} = 0{,}779x - 61{,}798$$

Als we nu voor x bijvoorbeeld een lengte van 195 cm invullen, vinden we als geschat gewicht:

$$\hat{y} = 0{,}779 \times 195 - 61{,}798 = 90{,}107$$

2.2 Spreiding van y rondom $\hat{y}$

De regressielijn wordt een betere schatter van de y-waarden naarmate de werkelijke y-waarden dichter bij de lijn liggen.

Om de betrouwbaarheid van de regressievergelijking te meten, berekenen we de variabiliteit van de werkelijke y-waarden rondom de regressielijn met volgende spreidingsformules:

$$s_e^2 = \frac{\Sigma\,(y - \hat{y})^2}{n - 2}$$

$$s_e = \sqrt{\frac{\Sigma\,(y - \hat{y})^2}{n - 2}}$$

s_e, schatter van σ_ε, is de steekproefstandaardafwijking van de afwijkingen e gelijk aan $y - \hat{y}$. Hoe kleiner s_e is, des te kleiner de spreiding van de y-waarden rond de regressielijn is.

In Figuur 93 hierna hebben we s_e berekend voor ons voorbeeld.

	A	B	C	D	E	F	G
1							
2		X	Y	$\hat{Y}$	Y-$\hat{Y}$	(Y-$\hat{Y}$)^2	
3		165	71	66,74	4,26	18,17	
4		170	70	70,63	-0,63	0,40	
5		180	79	78,42	0,58	0,33	
6		168	71	69,07	1,93	3,71	
7		184	80	81,54	-1,54	2,37	
8		168	67	69,07	-2,07	4,30	
9		184	79	81,54	-2,54	6,44	
10		160	58	62,84	-4,84	23,44	
11		182	88	79,98	8,02	64,32	
12		190	84	86,21	-2,21	4,89	
13		187	82	83,88	-1,88	3,52	
14		175	76	74,53	1,47	2,17	
15							
16	Σ(Y-$\hat{Y}$)^2	134,07					
17	Σ(Y-$\hat{Y}$)^2/n-2	13,41					
18	s_e	3,66					
19							

Figuur 93 Berekening van de steekproefstandaardafwijking s_e

2.3 Hypothesetoetsen voor de populatieparameters van de populatie-regressievergelijking

De regressievergelijking $\hat{y} = ax + b$ is een schatter van de populatieregressievergelijking $\hat{y} = AX + B$. A is de populatierichtingscoëfficiënt en B is de coëfficiënt die aangeeft welk stuk van de y-as door de populatieregressielijn wordt afgesneden.

Ook hier is het zo dat de werkelijke waarden van y meestal niet op de populatieregressielijn zullen liggen. Er zijn afwijkingen die we aanduiden met het symbool ε (epsilon). De standaardafwijking van deze storingsterm is σ_ε, waarvan s_e de schatter is. Dus σ_ε staat voor de afwijkingen van y t.o.v. $\hat{y}$ in de populatie en s_e staat voor de afwijking van y ten opzichte van $\hat{y}$ in de steekproef.
Deze afwijkingen kunnen bijvoorbeeld ontstaan doordat niet alle variabelen die een invloed uitoefenen op y gekend zijn of doordat de relatie niet zuiver lineair is.

2.3.1 Hypothesetoets voor ε

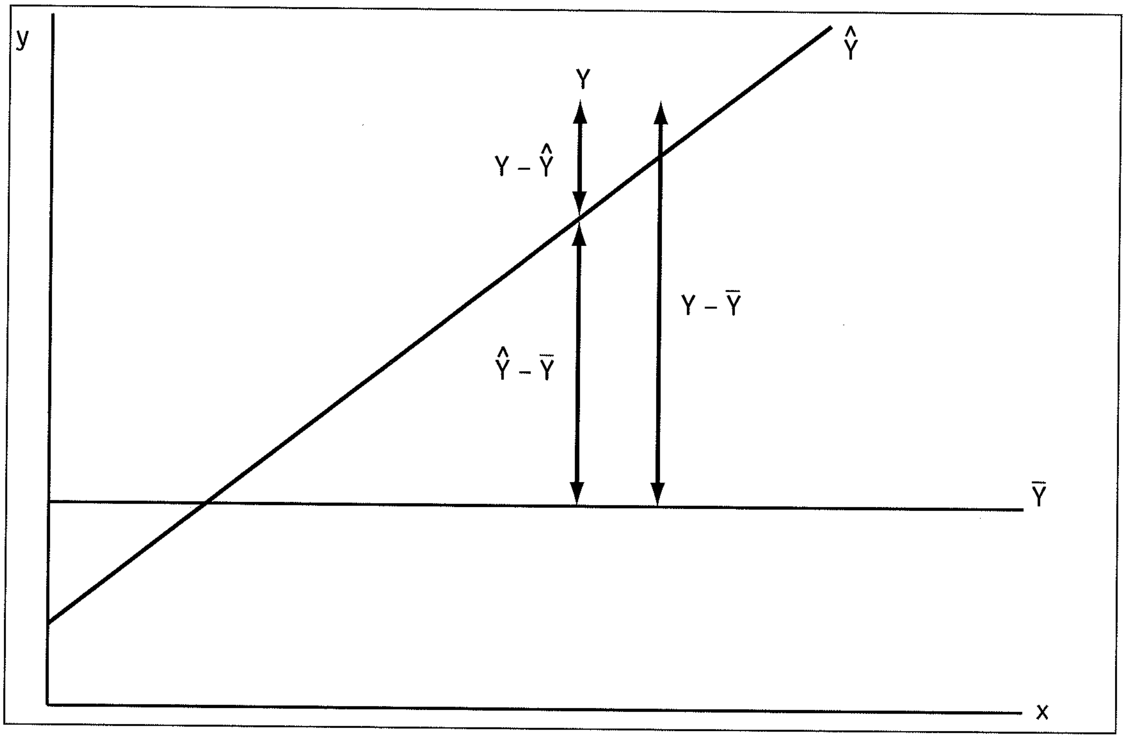
De hypothesetoets voor ε voeren we uit om na te gaan hoe goed de regressielijn is aangepast aan het spreidingsdiagram. Deze test moet ons dus toelaten te besluiten of onze regressielijn werkelijk een verband bewijst tussen de beschouwde variabelen, of niet.

De hypothesen voor deze test zijn:

H_0: geen relatie tussen y en x

H_1: H_0 wordt verworpen

Om de test uit te voeren, moeten we eerst een aantal afwijkingen bekijken. Figuur 94 beschouwt één punt y uit een spreidingsdiagram en maakt duidelijk dat de afwijking van het punt y t.o.v. zijn gemiddelde waarde $\bar{y}$ bestaat uit twee afwijkingen: de afwijking van y t.o.v. $\hat{y}$ en de afwijking van $\hat{y}$ t.o.v. $\bar{y}$ of: $y - \bar{y} = (y - \hat{y}) + (\hat{y} - \bar{y})$.



Figuur 94 De verschillende afwijkingen van één punt y

De afwijkingen leveren de volgende varianties op:

> variantie van e $= \dfrac{\Sigma\,(y - \hat{y})^2}{n - 2}$

Deze variantie geeft ons een idee van de variabiliteit rond de regressielijn. Dit wordt ook de gemiddelde kwadratensom van de residuen genoemd (Engels: Mean Square Residual) of de niet-verklaarde variantie.

> variantie van de regressielijn $= \dfrac{\Sigma\,(\hat{y} - \bar{y})^2}{1}$

De variantie van de regressielijn is de variantie als gevolg van afwijkingen van de regressielijn rond de verwachte waarde $\bar{y}$. Een andere term voor deze variantie is de gemiddelde kwadratensom van de regressie (Engels: Mean Square Regression) of de verklaarde variantie door de regressielijn.

> variantie van y $= \dfrac{\Sigma\,(y - \bar{y})^2}{n - 1}$

De variantie van y staat voor de totale variantie en wordt ook aangeduid met de term gemiddelde kwadratensom van de totale variatie (Engels: Mean Square Total).

Op basis van deze variantieanalyse kunnen we nu stellen dat er sprake zal zijn van een lineair verband als de variantie van de regressielijn of de verklaarde variantie groter is dan de variantie van ε of de niet verklaarde variantie.
Dus herformuleren we onze hypothesen als volgt:

H_0: $\sigma^2_{\text{regressie}} < \sigma^2_\varepsilon$
H_1: $\sigma^2_{\text{regressie}} > \sigma^2_\varepsilon$

We weten dat we, om varianties te testen, gebruikmaken van de F-ratio. Die wordt nu geformuleerd als:

$$F = \frac{\text{verklaarde variantie}}{\text{niet-verklaarde variantie}} = \frac{\text{variantie van de regressielijn}}{\text{variantie van e}}$$

We passen dit toe op ons voorbeeld en berekenen eerst de teller en noemer van F:
> $s^2_e = 13{,}41$ (zie Figuur 93)
> variantie van de regressielijn = $\Sigma\ (\hat{y} - \bar{y})^2 / 1$
$= (66{,}74 - 75{,}42)^2 + (70{,}63 - 75{,}42)^2 + (78{,}42 - 75{,}42)^2 + (69{,}07 - 75{,}42)^2 + (81{,}54 - 75{,}42)^2 + (69{,}07 - 75{,}42)^2 + (81{,}54 - 75{,}42)^2 + (62{,}84 - 75{,}42)^2 + (79{,}98 - 75{,}42)^2 + (86{,}21 - 75{,}42)^2 + (83{,}88 - 75{,}42)^2 + (74{,}53 - 75{,}42)^2$
$= 75{,}34 + 22{,}94 + 9 + 40{,}32 + 37{,}45 + 40{,}32 + 37{,}45 + 158{,}26 + 20{,}79 + 116{,}42 + 71{,}57 + 0{,}79$
$= 630{,}65$

Dat geeft voor F:

$$F = \frac{630{,}65}{13{,}41} = 47{,}03$$

Met de Excelfunctie F. INV. RECHTS berekenen we de kritieke waarde van F, wetende dat het aantal vrijheidsgraden voor de teller 1 is en voor de noemer (12 – 2) of 10, en $\alpha = 0{,}05$. We vinden voor F_K een waarde gelijk aan 4,96. Omdat de berekende F-waarde groter is dan F_K, verwerpen we H_0 en besluiten dat de regressielijn goed aangepast is aan het spreidingsdiagram en er dus sprake is van een significant lineair verband tussen de lengte en het gewicht van de studenten.

2.3.2 Hypothesetest voor A

Er zijn twee mogelijkheden om A te testen:
> De verwachte relatie is positief, dan moet aangetoond worden dat de richtingscoëfficiënt groter is dan nul. De hypothesen zijn dan:

H_0: $A \leq 0$
H_1: $A > 0$

> De verwachte relatie is negatief, dan moet bewezen worden dat de richtingscoëfficiënt kleiner is dan nul. De hypothesen zijn:

H_0: $A \geq 0$
H_1: $A < 0$

De toetsingsgrootheid is t:

$$t = \frac{a - A}{s_a}$$

s_a is de steekproefstandaardafwijking voor a en de formule ervan is:

$$s_a = \frac{s_e}{s_x \sqrt{n - 1}}$$

In Figuur 95 passen we de t-test toe op ons voorbeeld.

B23 =T.INV(B21;10)

	A	B	C	D	E	F
1		X(lengte)				
2		165				
3		170				
4		180				
5		168				
6		184				
7		168				
8		184				
9		160				
10		182				
11		190				
12		187				
13		175				
14						
15	s_X	9,718				
16	s_e	3,66				
17	s_a	0,1134				
18	a	0,779				
19	t	6,86				
20	α	0,05				
21	kans p	0,95				
22	v = n - 2	10				
23	t_k	1,81				
24						

Figuur 95 Hypothesetoets voor de richtingscoëfficiënt A

Met de functie STDEV berekenen we de standaarddeviatie van X en vinden zo voor s_a een waarde van 0,11. De t-test levert een t-waarde op van 6,86, die we moeten vergelijken met t_k en die we hebben gevonden met de functie T. INV.
Omdat t groter is dan t_k, kunnen we H_0 verwerpen en stellen dat de positieve samenhang beduidend is of m.a.w. verwacht mag worden.

2.3.3 Hypothesetoets voor B

Bij het uitvoeren van deze toets moet aangetoond worden dat B verschillend is van nul. De hypothesen zijn:
H_0: B = 0
H_1: B ≠ 0

Opnieuw gebruiken we de t-test. Voor het testen van B is de formule voor t:

$$t = \frac{b - B}{s_b}$$

s_b is de steekproefstandaardafwijking voor b met als formule:

$$s_b = s_e \cdot \sqrt{\frac{1}{n} + \frac{(\bar{x})^2}{\Sigma(x_i - \bar{x})^2}}$$

Als we s_b uitrekenen voor ons voorbeeld, vinden we:

$$s_b = 3{,}66 \,.\, \sqrt{\frac{1}{12} + \frac{(176{,}08)^2}{1038{,}92}}$$

$$= 20{,}023$$

Als we t invullen:

$$t = \frac{-61{,}798 - 0}{20{,}023} = -3{,}086$$

dan kunnen we met v = n - 2 of 10 en $\alpha = 0{,}05$ de berekende t-waarde vergelijken met t_k voor een tweezijdige test, omdat H_1 stelt dat de regressieconstante B verschillend is van nul.

Met de functie T. INV vinden we voor t_k een waarde van −2,23. Omdat t in de linkerstaart negatiever is dan t_k, verwerpen we H_0.

2.4 Werkbladfuncties

Voor regressieanalyse beschikt Excel over een aantal werkbladfuncties en matrixfuncties die het rekenwerk volledig voor hun rekening nemen.

In Figuur 96 hernemen we de gegevens van de 12 studenten waarvan we de gewichten Y en de lengte X hebben opgetekend.

	A	B	C	D
1		X	Y	
2	1	165	71	
3	2	170	70	
4	3	180	79	
5	4	168	71	
6	5	184	80	
7	6	168	67	
8	7	184	79	
9	8	160	58	
10	9	182	88	
11	10	190	84	
12	11	187	82	
13	12	175	76	
14				

Figuur 96 Gegevens gebruikt om de regressiewerkbladfuncties te illustreren

2.4.1 RICHTING, SNIJPUNT en STAND. FOUT. YX (SLOPE, INTERCEPT, STEYX)

Voor het gebruik van deze drie functies werken we als volgt:

> in het werkblad hebben we de gegevens ingevoerd en selecteren we een cel, bv. B16;
> we klikken op de f_x-knop en openen het dialoogvenster 'Functie invoegen';
> we selecteren de functie 'RICHTING' (SLOPE) om de richtingscoëfficiënt te berekenen van de regressielijn;
> vervolgens vullen we in het vak 'Y-bekend' van het dialoogvenster 'Functieargumenten' de kolom in met de y-waarden of het bereik C2:C13;
> in het vak 'X-bekend' brengen we de kolom in met x-waarden of het bereik B2:B13;
> we sluiten het venster 'Functieargumenten' en het antwoord voor de richtingscoëfficiënt verschijnt in cel B16: +0,779 afgerond;
> we herhalen deze procedure voor de functies SNIJPUNT (INTERCEPT) en STAND. FOUT. YX (STEYX) en krijgen de volgende resultaten: cel B17 toont dat we met de functie SNIJPUNT voor b, de regressieconstante, een waarde −61,798 vinden en in cel B18 verschijnt voor s_e, de steekproefstandaardafwijking van e, de waarde 3,661 berekend met de functie STAND. FOUT. YX.

In Figuur 97 tonen we de resultaten van de drie gebruikte functies in een Excelwerkblad.

	A	B	C	D
1		X	Y	
2	1	165	71	
3	2	170	70	
4	3	180	79	
5	4	168	71	
6	5	184	80	
7	6	168	67	
8	7	184	79	
9	8	160	58	
10	9	182	88	
11	10	190	84	
12	11	187	82	
13	12	175	76	
14				
15				
16	a	0,779		
17	b	-61,798		
18	s_e	3,661		
19				

Figuur 97 Berekening van de regressiecoëfficiënt, de constante en s_e, respectievelijk met de functies RICHTING, SNIJPUNT en STAND. FOUT. YX

2.4.2 VOORSPELLEN.LINEAIR (FORECAST.LINEAR)

Met de functie VOORSPELLEN kunnen we voor een gekozen waarde van x een waarde van y voorspellen, aan de hand van de lineaire regressierelatie tussen X en Y.

Hier volgen de stappen voor deze functie:

> in het werkblad van Figuur 98 voer je de gegevens x en y van de 12 studenten opnieuw in, respectievelijk in de bereiken B3:B14 en C3:C14;
> cel B17 selecteer je om de voorspelde waarde $\hat{y}$ te tonen voor de x-waarde gelijk aan 188 cm;
> je klikt op de f_x-knop om het venster 'Functie invoegen' te openen, je selecteert de statistische functie 'VOORSPELLEN.LINEAIR' en klikt op 'OK' om het venster 'Functie-argumenten' te openen;
> in het X-vak voer je de waarde 188 in, waarvoor we de voorspelde waarde van y willen kennen;
> in het Y-bekend-vak voer je het bereik C3:C14 in en in het vak X-bekend zet je het bereik B3:B14;
> je klikt op 'OK' en in cel B17 van het werkblad lezen we de voorspelde waarde 84,70 kg af.

B17 =VOORSPELLEN(188;C3:C14;B3:B14)

	A	B	C	D	E	F	G
1							
2	Student	X	Y				
3	1	165	71				
4	2	170	70				
5	3	180	79				
6	4	168	71				
7	5	184	80				
8	6	168	67				
9	7	184	79				
10	8	160	58				
11	9	182	88				
12	10	190	84				
13	11	187	82				
14	12	175	76				
15							
16	X	188					
17	Y	84,70					
18							

Figuur 98 Berekening van $\hat{y}$ met de functie VOORSPELLEN

2.4.3 TREND (TREND)

Met de Excelfunctie TREND kunnen we een tweetal berekeningen uitvoeren:

a Berekening van alle geschatte waarden van y voor de x-waarden uit onze steekproef

Hier volgen de uit te voeren stappen:

> in Figuur 99 voeren we opnieuw de x- en y-gegevens in en dat in de bereiken B3:B14 en C3:C14;
> we selecteren een kolom voor de voorspelde y-waarden, zijnde het bereik D3:D14;
> we klikken op de f_x-knop, kiezen in het venster 'Functie invoegen' de TREND-functie en openen het venster 'Functieargumenten';
> in het vak Y-bekend voeren we het bereik C3:C14 in en in het vak X-bekend het bereik B3:B14;
> in het vak Const voeren we de logische waarde 'WAAR' in om de b-constante te berekenen;
> we drukken op Ctrl + Shift + OK om het resultaat te tonen in het bereik D3:D14.

D3 | {=TREND(C3:C14;B3:B14;;WAAR)}

	A	B	C	D	E	F
1						
2	Student	X	Y	Ŷ		
3	1	165	71	66,7798989		
4	2	170	70	70,6761851		
5	3	180	79	78,4687575		
6	4	168	71	69,1176707		
7	5	184	80	81,5857865		
8	6	168	67	69,1176707		
9	7	184	79	81,5857865		
10	8	160	58	62,8836127		
11	9	182	88	80,027272		
12	10	190	84	86,2613299		
13	11	187	82	83,9235582		
14	12	175	76	74,5724713		
15						

Figuur 99 Voorspelde y-waarden voor de x-waarden uit de steekproef met de TREND-functie

b Berekening van geschatte y-waarden voor nieuwe x-waarden

In Figuur 100 hebben we de x- en y-gegevens in de kolommen B en C ingevoerd om de tweede toepassing van TREND te illustreren:

- in kolom D hebben we het gegevensbereik D3:D14 ingevoerd dat nieuwe x-waarden bevat;
- we selecteren het gegevensbereik E3:E14 voor het resultaat van TREND;
- we klikken op de f_x-knop, selecteren de TREND-functie in het venster 'Functie invoegen' en openen het venster 'Functieargumenten';
- in het vak Y-bekend voeren we het bereik C3:C14 in en in het vak X-bekend het bereik B3:B14;
- in het vak X-nieuw voeren we het gegevensbereik D3-D14 in;
- in het vak Const. vullen we de logische waarde 'WAAR' in om de b-constante te berekenen;
- omdat TREND een matrixfunctie is, drukken we de combinatie Ctrl + Shift + OK in om het resultaat te tonen in het bereik E3:E14.

E3 {=TREND(C3:C14;B3:B14;D3:D14;WAAR)}

	A	B	C	D	E	F
1						
2	Student	X	Y	Nieuwe X-waarden	$\hat{Y}$ voor nieuwe X-waarden	
3	1	165	71	163	65,22138445	
4	2	170	70	169	69,89692789	
5	3	180	79	181	79,24801476	
6	4	168	71	183	80,80652924	
7	5	184	80	185	82,36504372	
8	6	168	67	167	68,33841341	
9	7	184	79	186	83,14430095	
10	8	160	58	159	62,1043555	
11	9	182	88	189	85,48207267	
12	10	190	84	191	87,04058715	
13	11	187	82	192	87,81984439	
14	12	175	76	173	73,01395685	
15						

Figuur 100 Geschatte y-waarden voor nieuwe x-waarden met de TREND-functie

2.4.4 LIJNSCH (LINEST)

De functie LIJNSCH geeft als resultaat een matrix die een rechte lijn beschrijft die het best past bij de gegevens.

Figuur 101 toont de x- en y-steekproefgegevens die nodig zijn voor de uitvoering van deze functie.
We doen het volgende:

> we selecteren een matrix die vijf rijen en twee kolommen heeft of het bereik E2:F6;
> we klikken op de f_x-knop, selecteren in het venster 'Functie invoegen' de functie LIJNSCH en klikken op 'OK' om het venster 'Functieargumenten' te openen;
> in het vak Y-bekend voeren we het Y-bereik C2:C13 in;
> in het vak X-bekend voeren we het X-bereik B2:B13 in;
> in het vak Const vullen we de logische waarde WAAR in om de b-constante te berekenen;
> in het vak Stat tikken we de logische waarde 'WAAR' in, die als resultaat een aantal regressiegrootheden zal opleveren naast de regressierichtingscoëfficiënt en de b-constante;
> we drukken op de combinatie Ctrl + Shift + OK en verkrijgen de resultaten die de regressielijn beschrijven in het bereik E2:F6.

De functie geeft niet aan wat de berekende resultaten eigenlijk zijn. Daarom hebben we zelf de omschrijvingen in het werkblad toegevoegd.

E2 | {=LIJNSCH(C2:C13;B2:B13;WAAR;WAAR)}

	A	B	C	D	E	F	G	H
1		X	Y					
2	1	165	71	a	0,779257239	-61,79754552	b	
3	2	170	70	s_a	0,113587804	20,02882398	s_b	
4	3	180	79	R kwadraat	0,824761311	3,661188243	s_e	
5	4	168	71	F	47,06502419	10	v	
6	5	184	80	Som van de gekwadrateerde variaties van de regressielijn	630,8736732	134,0429935	Som van de gekwadrateerde variaties van e	
7	6	168	67					
8	7	184	79					
9	8	160	58					
10	9	182	88					
11	10	190	84					
12	11	187	82					
13	12	175	76					
14								

Figuur 101 De LIJNSCH-functie

2.5 Gegevensanalysefunctie: Regressie (Regression)

In Figuur 102a tonen we opnieuw de steekproefgegevens van de lengten en gewichten van de twaalf studenten, waarop we de gegevensanalysefunctie Regressie zullen toepassen.

We gebruiken de volgende werkwijze:

> we tikken onze steekproefgegevens in het Excelwerkblad in;
> in het lint selecteren we het tabblad 'Gegevens', dan klikken we op de knop 'Gegevensanalyse' waardoor het dialoogvenster 'Gegevensanalyse' wordt geopend. In dit venster kiezen we de functie 'Regressie' en klikken op 'OK' om de functie te openen;
> in het vak 'Invoerbereik Y' voeren we het gegevensbereik D2:D13 in;
> in het vak 'Invoerbereik X' voeren we het gegevensbereik C2:C13 in;
> we selecteren de optie 'Betrouwbaarheidsniveau' en geven een standaardniveau op van 95 % die we op de regressie willen toepassen;
> als uitvoeroptie selecteren we 'Uitvoerbereik' en voeren cel A14 in door in het werkblad te klikken op cel A14. Deze cel verwijst naar de linkerbovenhoek van de uitvoertabel;
> bij de 'Storingen' selecteren we alle opties. Die laten ons de afwijkingen zien tussen de y- en $\hat{y}$-waarden;
> ten slotte selecteren we ook nog 'Grafiek met kanswaarden' om een grafiek te maken met de normale kansverdeling;
> we klikken op 'OK' en de uitvoertabel verschijnt in het werkblad van Figuur 102a. De geproduceerde grafieken tonen we afzonderlijk.

De eerste drie rijen van de tabel *Gegevens voor de regressie* bevatten gegevens van grootheden die we bespreken in het volgende hoofdstuk over correlatie.

	A	B	C	D	E	F	G	H	I
1		Student	X	Y					
2		1	165	71					
3		2	170	70					
4		3	180	79					
5		4	168	71					
6		5	184	80					
7		6	168	67					
8		7	184	79					
9		8	160	58					
10		9	182	88					
11		10	190	84					
12		11	187	82					
13		12	175	76					
14	SAMENVATTING UITVOER								
15									
16	*Gegevens voor de regressie*								
17	Meervoudige correlatiecoëfficiënt R	0,908163703							
18	R-kwadraat	0,824761311							
19	Aangepaste kleinste kwadraat	0,807237443							
20	Standaardfout	3,661188243							
21	Waarnemingen	12							
22									
23	Variantie-analyse								
24		*Vrijheidsgraden*	*Kwadratensom*	*Gemiddelde kwadraten*	*F*	*Significantie F*			
25	Regressie	1	630,8736732	630,8736732	47,065	4,40212E-05			
26	Storing	10	134,0429935	13,40429935					
27	Totaal	11	764,9166667						
28									
29		*Coëfficiënten*	*Standaardfout*	*T- statistische gegevens*	*P-waarde*	*Laagste 95%*	*Hoogste 95%*	*Laagste 95,0%*	*Hoogste 95,0%*
30	Snijpunt	-61,79754552	20,02882398	-3,085430556	0,01153	-106,4245464	-17,17054465	-106,4245464	-17,17054465
31	Variabele X 1	0,779257239	0,113587804	6,860395338	4,4E-05	0,526167841	1,032346638	0,526167841	1,032346638
32									
33									
34									
35	STORINGEN					KANS			
36									
37	*Waarneming*	*Voorspelde Y*	*Storingen*	*Standaardstoringen*		*Percentiel*	*Y*		
38	1	66,77989893	4,220101067	1,208918811		4,166666667	58		
39	2	70,67618513	-0,676185129	-0,193704584		12,5	67		
40	3	78,46875752	0,53124248	0,152183329		20,83333333	70		
41	4	69,11767065	1,882329349	0,539224849		29,16666667	71		
42	5	81,58578648	-1,585786476	-0,454275163		37,5	71		
43	6	69,11767065	-2,117670651	-0,606642316		45,83333333	76		
44	7	81,58578648	-2,585786476	-0,740741955		54,16666667	79		
45	8	62,88361274	-4,883612738	-1,398992871		62,5	79		
46	9	80,027272	7,972728002	2,283921808		70,83333333	80		
47	10	86,26132991	-2,261329911	-0,647795924		79,16666667	82		
48	11	83,92355819	-1,923558194	-0,551035544		87,5	84		
49	12	74,57247132	1,427528676	0,408939559		95,83333333	88		
50									

Figuur 102a Uitvoertabel van de gegevensanalysefunctie Regressie

De vierde rij toont s_e; de laatste rij geeft het aantal steekproefelementen waarop de regressie wordt toegepast.

De *Variantieanalysetabel* geeft de waarde op van de variantie van de regressielijn en van s_e^2. Met deze waarden wordt F dan berekend. In de *F-Significantiekolom* vinden we een waarde kleiner dan $\alpha = 0{,}05$, wat een bewijs is voor de verwerping van de H_0-hypothese die stelt dat $\sigma^2_{regressie} < \sigma^2_\varepsilon$. Zo kunnen we opnieuw besluiten dat er een lineair verband is tussen de lengte en het gewicht van de studenten.

Onder de *Variantieanalysetabel* vinden we een tabel met gegevens over de regressiecoëfficiënten.

In de kolom *Coëfficiënten* staan de waarden voor a en b. De richtingscoëfficiënt is aangeduid met de naam Variabele X 1. De kolom *Standaardfout* toont de waarden van s_b en s_a.

In de kolom *T-statistische gegevens* vinden we de berekende t-waarden voor het testen van B en A.

De kolom met de P-waarden laat ons toe uit te maken of we H_0 van iedere test al of niet kunnen verwerpen. Zowel voor de B- als voor de A-test is de verwerping van H_0 het geval. De resterende kolommen geven nog meer resultaten rond de t-testen van A en B.

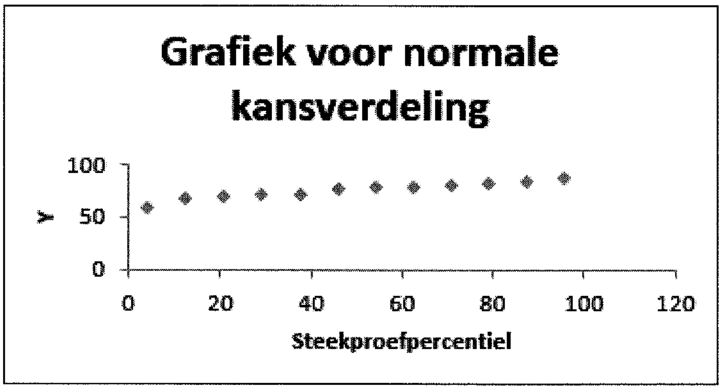
Een tabel *STORINGEN* levert een kolom op met $\hat{y}$-waarden en storingswaarden voor ieder steekproefelement. In de laatste kolom wordt nog voor iedere geobserveerde waarde de standaardstoring berekend die gevonden wordt met de formule:

$$\text{standaardstoring} = \frac{\text{storing} - \text{gemiddelde storing}}{s_e}$$

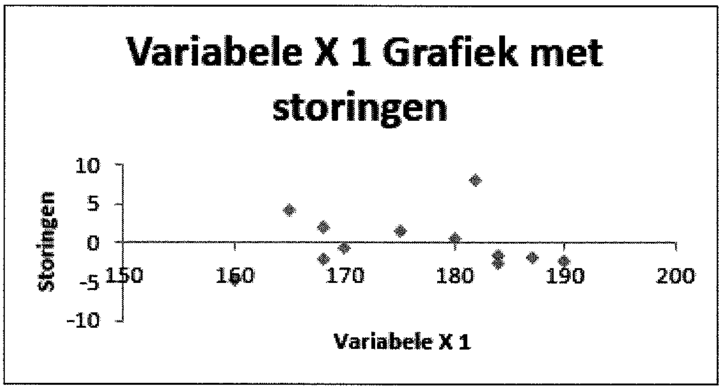
Deze tabel is interessant om de variabiliteit rondom de regressielijn te analyseren.

In de laatste tabel, *KANS*, worden voor de y-waarden van de steekproef de overeenstemmende percentielen berekend.

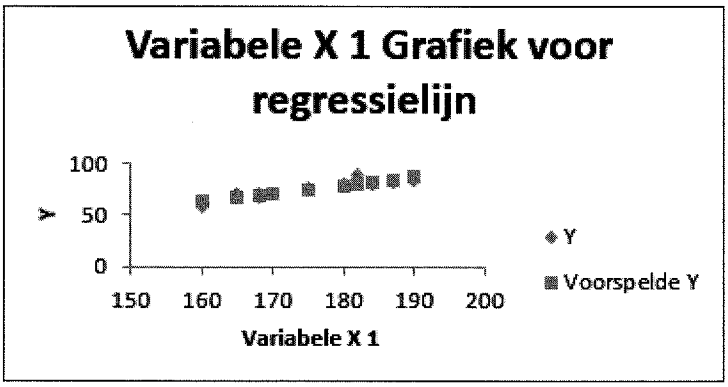
Ten slotte is er nog de grafische uitvoer van de Regressiefunctie, die bestaat uit drie grafieken:



Figuur 102b Grafiek voor normale kansverdeling geeft de grafische weergave van de KANS-tabel



Figuur 102c Variabele x1 Grafiek met storingen, toont voor iedere x-waarde in de steekproef de corresponde-rende storing



Figuur 102d Variabele x1 Grafiek voor regressielijn brengt y- en $\hat{y}$-waarden in beeld voor alle x-waarden uit de steekproef

3 Meervoudige regressie

Als we meer dan één onafhankelijke variabele gebruiken om de afhankelijke variabele y te schatten, dan spreken we van meervoudige regressie. Met andere woorden, we gaan meer informatie gebruiken om y te schatten.

De schattingsvergelijking met bijvoorbeeld twee onafhankelijke variabelen ziet er als volgt uit:

$$\hat{y} = a_1x_1 + a_2x_2 + b$$

De meetkundige figuur van deze schattingsvergelijking is een plat vlak. Zoals bij de lineaire regressie zullen er y-waarden zijn die boven het vlak liggen en andere die onder het vlak zullen liggen. Het zal erop aankomen een vlak te vinden dat de som van de kwadraten van de afwijkingen t.o.v. de werkelijke y-waarden minimaliseert.

Als voorbeeld nemen we een onderneming die tv-toestellen verkoopt. De manager meent dat de twee voornaamste determinanten om de verkoop van zijn toestellen te verklaren, zijn: het aantal advertenties en de prijs. Daarover verzamelde hij de volgende informatie:

Omzet (aantal eenheden)	Advertenties (aantal advertenties)	Prijs (€)
66	3	1 250
121	6	1 150
140	10	1 400
164	13	1 300
34	9	1 450
48	6	1 400
70	4	1 350
100	5	1 130
185	9	1 100
90	4	1 200

De statistische formules voor de berekening van de coëfficiënten laten we hier achterwege en we gebruiken onmiddellijk de Excelfuncties.

3.1 Werkbladfuncties

In Figuur 103 voeren we in het werkblad de omzet-, advertentie- en prijswaarden in, en dat achtereenvolgens in de kolommen B, C en D. Op deze invoer passen we de Excelfuncties TREND en LIJNSCH toe.

3.1.1 TREND

Met deze functie kunnen we voor de waarden van de onafhankelijke variabelen, geschatte waarden van y vinden.

De procedure ziet er als volgt uit:

> je selecteert in kolom E het bereik E2:E11 voor de geschatte of voorspelde waarden van O (= omzet);
> je klikt op de f_x-knop, opent het venster 'Functie invoegen', selecteert de functie 'TREND', klikt op 'OK' en opent het venster 'Functieargumenten';
> in het vak Y-bekend voer je het bereik B2:B11 in en in het vak X-bekend het bereik C2:D11;
> in het vak Const voer je de logische waarde 'WAAR' in voor de berekening van de constante b;
> je drukt op de combinatie Ctrl + Shift + OK en de resultaten verschijnen in de geselecteerde kolom E2:E11.

E2 | {=TREND(B2:B11;C2:D11;;WAAR)}

	A	B	C	D	E	F	G
1		O	A	P	Ô		
2		66	3	1250	62,5623308		
3		121	6	1150	126,222951		
4		140	10	1400	102,063892		
5		164	13	1300	165,724512		
6		34	9	1450	76,102821		
7		48	6	1400	55,1099821		
8		70	4	1350	45,8556208		
9		100	5	1130	120,173511		
10		185	9	1100	175,660977		
11		90	4	1200	88,523402		
12							

Figuur 103 De geschatte omzetwaarden $\hat{O}$ met de functie TREND

3.1.2 LIJNSCH

Met deze functie berekenen we, zoals gezien, de regressiecoëfficiënten. De volgende stappen moeten uitgevoerd worden:

> nadat we de steekproefgegevens hebben ingevoerd in het werkblad van Figuur 104, selecteren we een matrix met drie kolommen en vijf rijen, zijnde B14:D18;
> we klikken op de f_x-knop, selecteren in het venster 'Functie invoegen' de functie 'LIJNSCH', klikken op 'OK' en het venster 'Functieargumenten' wordt geopend;

- in het vak Y-bekend voeren we het bereik B2:B11 in voor de O-waarden;
- in het vak X-bekend voeren we voor de variabelen A (= advertenties) en P (= prijs) het bereik C2:D11 in;
- in het vak Const voeren we de logische waarde 'WAAR' in voor de berekening van de b-constante;
- in het vak Stat voeren we ook de logische waarde 'WAAR' in voor de berekening van een aantal statistische grootheden;
- we drukken op de combinatie Ctrl + Shift + OK zodat we de resultaten van deze functie zien in het geselecteerde bereik B14:D18.

De omschrijvingen van de regressiegrootheden hebben we in het werkblad toegevoegd omdat de functie zelf geen omschrijvingen geeft.

In de eerste rij vinden we de twee coëfficiënten voor de X-variabelen P en A en de coëfficiënt die overeenstemt met b. De tweede rij levert de waarden op van s_{a_2}, s_{a_1} en s_b.

Op basis van de gevonden regressiecoëfficiënten ziet onze regressievergelijking er dan als volgt uit:

$$\hat{y} = 11{,}738\ X_1 - 0{,}284\ X_2 + 382{,}912$$

of aangepast aan de benamingen van onze variabelen:

$$\hat{O} = 11{,}738\ A - 0{,}284\ P + 382{,}912$$

De middelste rij toont R^2, een maat voor de sterkte van de samenhang tussen O en de onafhankelijke variabelen A en P, en s_e.

B14 | {=LIJNSCH(B2:B11;C2:D11;WAAR;WAAR)}

	A	B	C	D	E	F
1		O	A	P		
2		66	3	1250		
3		121	6	1150		
4		140	10	1400		
5		164	13	1300		
6		34	9	1450		
7		48	6	1400		
8		70	4	1350		
9		100	5	1130		
10		185	9	1100		
11		90	4	1200		
12						
13		a_2	a_1	b		
14	Coëfficiënt	-0,28445	11,73848	382,9117		
15	Steekproefstandaardafwijking	0,068713	2,682746	85,0968		
16	R-kwadraat	0,805561	25,02495	#N/B	s_e	
17	F	14,50053	7	#N/B	v	
18	Som van de gekwadrateerde variaties van de regressielijn	18161,86	4383,738	#N/B	Som van de gekwadrateerde variaties van de residuen	
19						

Figuur 104 Regressiegrootheden berekend met de functie LIJNSCH

De F-grootheid in de vierde rij test de hypothese of de grafiek van de regressievergelijking goed is aangepast aan het spreidingsdiagram. In deze rij is ook het aantal vrijheidsgraden voor de noemer van de F-grootheid opgegeven.

Het aantal vrijheidsgraden voor de teller is gelijk aan het aantal coëfficiënten min 1. Als we de functie F. INV. RECHTS gebruiken voor de berekening van F_K met 7 en 2 vrijheidsgraden en met α = 0,05, vinden we voor F_k een waarde gelijk aan 4,737 of een bewijs tegen H_0. De grafiek van de regressievergelijking is dus goed aangepast aan het spreidingsdiagram.

De laatste rij geeft ten slotte de sommen van de gekwadrateerde variaties van de regressielijn en de gekwadrateerde variaties van de residuen.

3.2 Gegevensanalysefunctie: Regressie

Figuur 105a bevat opnieuw de steekproefgegevens die we nodig hebben om de functie Regressie uit te voeren: in kolom B de O of omzetwaarden, in kolom C de A-waarden of het aantal advertenties en in kolom D de P- of prijswaarden.

Na deze invoer van gegevens voeren we de volgende stappen uit:

> in het lint selecteren we het tabblad 'Gegevens', daarna klikken we op de knop 'Gegevensanalyse' en het venster 'Gegevensanalyse' verschijnt. In het venster 'Gegevensanalyse' selecteren we de functie 'Regressie' en klikken op 'OK' om het venster 'Regressie' te openen

> in het vak Invoerbereik Y voeren we het gegevensbereik B1:B11 in of de O-waarden;

> in het vak Invoerbereik X voeren we het gegevensbereik C1:D11 in of de A- en P-waarden;

> vervolgens selecteren we de optie 'Labels' om aan te geven dat de celbereiken ook de kolomtitels bevatten;

> in het vak 'Betrouwbaarheidsniveau' behouden we het standaardniveau van 95 % dat we willen toepassen op de regressie;

> bij de Uitvoeropties kiezen we de optie 'Nieuw werkblad';

> we selecteren alle opties bij de 'Storingen'. Deze opties laten ons de afwijkingen zien tussen de werkelijke en geschatte punten;

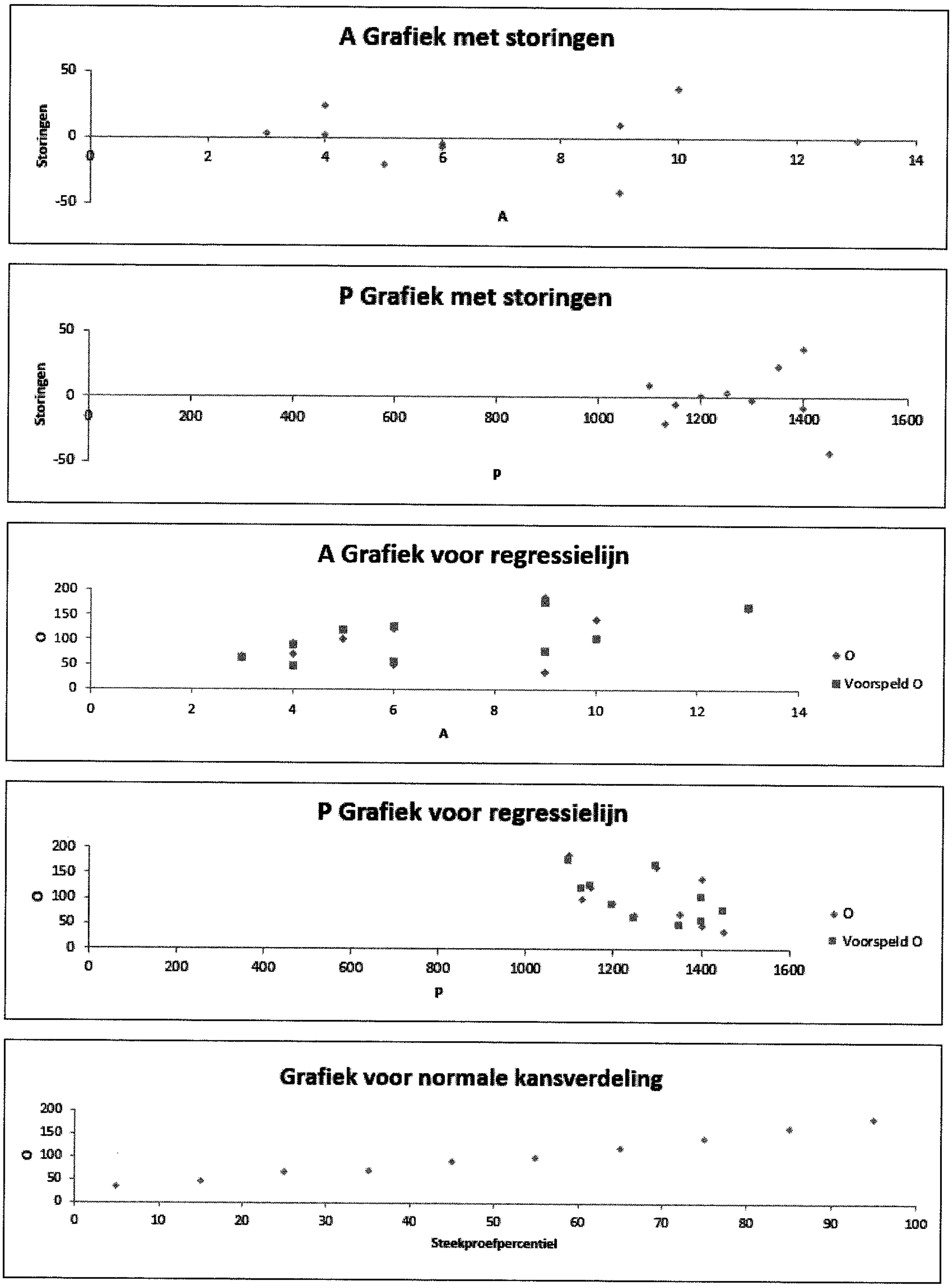
> we schakelen ook nog de optie 'Grafiek met kanswaarden' in die een grafiek maakt met de percentielen van de Y-veranderlijke;

> we klikken op 'OK' en de uitvoer verschijnt in Figuur 105b.

	A	B	C	D	E
1		O	A	P	
2		66	3	1250	
3		121	6	1150	
4		140	10	1400	
5		164	13	1300	
6		34	9	1450	
7		48	6	1400	
8		70	4	1350	
9		100	5	1130	
10		185	9	1100	
11		90	4	1200	
12					

Figuur 105a Steekproefgegevens voor meervoudige regressieanalyse

	A	B	C	D	E	F	G	H	I
1	SAMENVATTING UITVOER								
2									
3	*Gegevens voor de regressie*								
4	Meervoudige correlatiecoëfficiër	0,897530656							
5	R-kwadraat	0,805561278							
6	Aangepaste kleinste kwadraat	0,750007357							
7	Standaardfout	25,02495227							
8	Waarnemingen	10							
9									
10	Variantie-analyse								
11		*Vrijheidsgraden*	*Kwadratensom*	*Gemiddelde kwadraten*	*F*	*Significantie F*			
12	Regressie	2	18161,86235	9080,931174	14,5005	0,003241452			
13	Storing	7	4383,737653	626,2482361					
14	Totaal	9	22545,6						
15									
16		*Coëfficiënten*	*Standaardfout*	*T- statistische gegevens*	*P-waarde*	*Laagste 95%*	*Hoogste 95%*	*Laagste 95,0%*	*Hoogste 95,0%*
17	Snijpunt	382,9117416	85,09679539	4,499719876	0,0028	181,6897955	584,1336877	181,6897955	584,1336877
18	A	11,73847752	2,682746377	4,375545009	0,00325	5,394790377	18,08216466	5,394790377	18,08216466
19	P	-0,284451875	0,068713254	-4,139694456	0,00435	-0,4469329	-0,121970849	-0,4469329	-0,121970849
20									
21									
22									
23	STORINGEN					KANS			
24									
25	*Waarneming*	*Voorspeld O*	*Storingen*	*Standaardstoringen*		*Percentiel*	*O*		
26	1	62,56233079	3,43766921	0,155762555		5	34		
27	2	126,2229508	-5,22295082	-0,23665458		15	48		
28	3	102,0638922	37,93610777	1,718904495		25	66		
29	4	165,7245123	-1,724512255	-0,078138535		35	70		
30	5	76,10282097	-42,10282097	-1,907700407		45	90		
31	6	55,10998215	-7,109982146	-0,322156937		55	100		
32	7	45,85562084	24,14437916	1,0939942		65	121		
33	8	120,1735108	-20,17351079	-0,914072118		75	140		
34	9	175,6609771	9,339022886	0,423155916		85	164		
35	10	88,52340205	1,476597955	0,06690541		95	185		



Figuur 105b Uitvoer van de gegevensanalysefunctie Regressie

De gegevens uit de eerste drie rijen van de eerste tabel *Gegevens voor de regressie* bespreken we in het hoofdstuk over correlatie. In de vierde rij vinden we de berekening van s_e en in de laatste rij het aantal waarnemingen.

De tabel *Variantieanalyse* geeft de variantie van de regressielijn en de variantie van de residuen. Op basis van beide varianties wordt F berekend en we stellen vast dat de uitkomst in de kolom *Significantie F* bewijs levert tegen H_0.

De derde tabel van de coëfficiënten geeft ons eerst de waarden van de coëfficiënten en vervolgens de waarden van s_b, s_{a_1}, s_{a_2}.

In de kolom *T-statistische gegevens* vinden we de berekende t-waarden voor het testen van B, A1 en A2. De P-waarden leveren een bewijs tegen H_0. De andere kolommen geven nog bijkomende resultaten i.v.m. de t-testen.

De voorlaatste tabel *Storingen* bevat een kolom met de voorspelde O-waarden en twee kolommen met storingswaarden.

Tabel *Kans* toont voor de O-waarden de corresponderende percentielberekening.

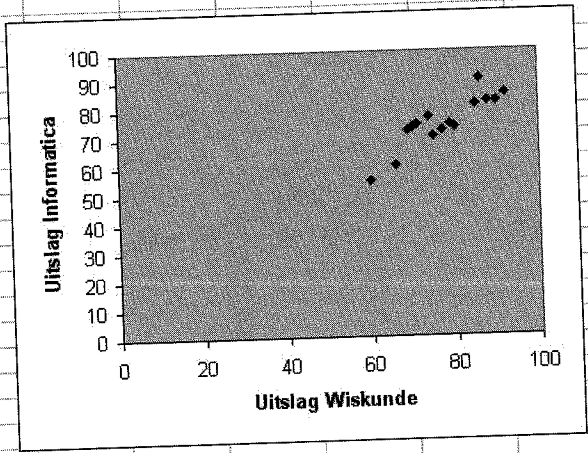
In de eerste twee grafieken worden respectievelijk tegenover de steekproefwaarden van de onafhankelijke variabelen A en P de storingswaarden uitgezet.
In de derde en vierde grafiek worden achtereenvolgens tegenover de A- en P-waarden de werkelijke en voorspelde O-waarden uitgezet. De uitvoer toont ten slotte nog een grafische weergave van de *Kans-tabel.*

hoofdstuk 18
Correlatie

1 Spreidingsdiagram

In Figuur 106 hebben we in het werkblad de uitslagen op 100 punten behaald op het vak wiskunde (= X) en het vak informatica (= Y) door 15 studenten weergegeven. Met de Excelprocedure voor 'Grafieken' hebben we een spreidingsdiagram getekend van de behaalde cijfers voor wiskunde en informatica.

Student	Wiskunde(= X)	Informatica (= Y)
1	70	73
2	79	74
3	85	81
4	71	74
5	88	82
6	75	70
7	90	82
8	66	60
9	86	90
10	92	85
11	74	77
12	60	55
13	69	72
14	80	73
15	77	72



Figuur 106 Spreidingsdiagram van de cijfers van 15 studenten voor wiskunde en informatica

2 Correlatie

Correlatie gaat de sterkte van samenhang na tussen twee variabelen.

De correlatie is in het voorbeeld van de studenten positief als met een toename van de uitslag voor wiskunde een toename van de uitslag voor informatica gepaard gaat.
We spreken van een negatieve correlatie als met een toename van de uitslag voor wiskunde een daling van de uitslag voor informatica gepaard gaat.

De correlatie r tussen twee variabelen wordt berekend met de volgende formule:

$$r = \frac{\frac{1}{n-1} \Sigma (x - \bar{x})(y - \bar{y})}{s_x \cdot s_y}$$

De teller berekent hoe x en y samen veranderen, of de covariantie.
Door de covariantie van X en Y te delen door het product van hun standaardafwijkingen maken we de uitkomst van de correlatie onafhankelijk van de eenheid waarin X en Y zijn uitgedrukt.

Een andere belangrijke eigenschap van r is dat zijn waarde begrensd is tussen –1 en +1.

Als X en Y perfect positief correleren, is r gelijk aan +1, als ze perfect negatief correleren is r gelijk aan –1. Als r gelijk is aan 0, is er geen correlatie tussen de variabelen.

In Figuur 107 hernemen we de gegevens en passen we de correlatieformule r toe. We berekenen achtereenvolgens $\bar{x}$, $\bar{y}$, s_x^2, s_y^2 en s_x en s_y met de Excelfuncties GEMIDDELDE, VAR. S en STDEV. S. In cel B22 berekenen we r en vinden als uitkomst 0,88, een sterke positieve correlatie.

B22 =(F17/14)/(B20*C20)

	A	B	C	D	E	F	G
1	Student	Wiskunde(= X)	Informatica (= Y)	$X - \bar{X}$	$Y - \bar{Y}$	$(X - \bar{X}) * (Y - \bar{Y})$	
2	1	70	73	-7,47	-1,67	12,44	
3	2	79	74	1,53	-0,67	-1,02	
4	3	85	81	7,53	6,33	47,71	
5	4	71	74	-6,47	-0,67	4,31	
6	5	88	82	10,53	7,33	77,24	
7	6	75	70	-2,47	-4,67	11,51	
8	7	90	82	12,53	7,33	91,91	
9	8	66	60	-11,47	-14,67	168,18	
10	9	86	90	8,53	15,33	130,84	
11	10	92	85	14,53	10,33	150,18	
12	11	74	77	-3,47	2,33	-8,09	
13	12	60	55	-17,47	-19,67	343,51	
14	13	69	72	-8,47	-2,67	22,58	
15	14	80	73	2,53	-1,67	-4,22	
16	15	77	72	-0,47	-2,67	1,24	
17						1048,33	
18	Gemiddelde	77,47	74,67				
19	Variantie	88,70	81,38				
20	Standaardafwijking	9,42	9,02				
21	n - 1	14					
22	r	0,88					
23							

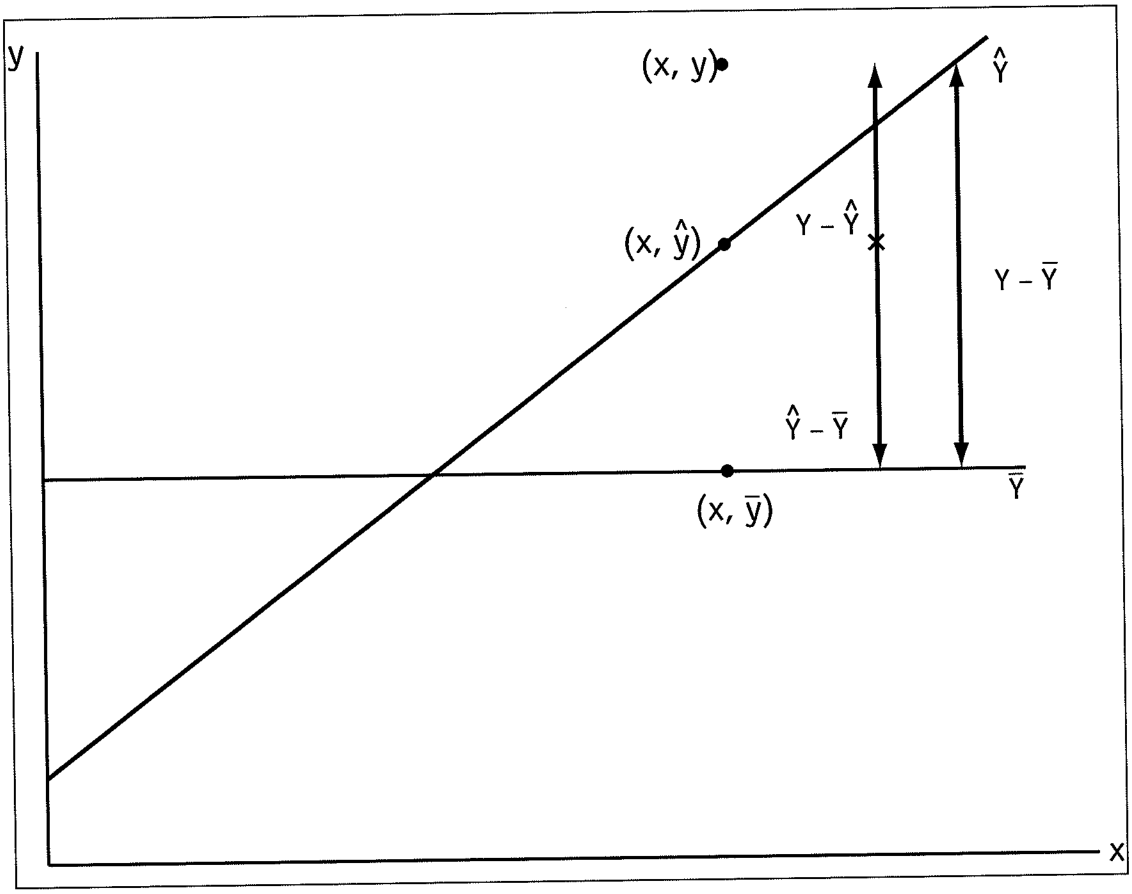
Figuur 107 Berekening van de correlatie r tussen uitslag wiskunde en uitslag informatica van studenten

3 De determinatiecoëfficiënt

In het vorige hoofdstuk hebben we gezien dat de afwijking van een punt y t.o.v. zijn gemiddelde waarde $\bar{y}$ bestaat uit twee afwijkingen: de afwijking van y t.o.v. $\hat{y}$ en de afwijking van $\hat{y}$ t.o.v. $\bar{y}$ of:

$$y - \bar{y} = (y - \hat{y}) + (\hat{y} - \bar{y})$$

In figuur 108 beelden we $y - \bar{y}$ opnieuw af voor een punt (x, y):



Figuur 108 Punt (x, y) in het spreidingsdiagram en zijn afstanden tot de regressielijn en zijn $\bar{y}$

Als we de variatie van de regressielijn delen door de totale variatie of de variatie van y, krijgen we een verhouding die we de determinatiecoëfficiënt r^2 noemen.

De formule is:

$$r^2 = \frac{\Sigma (\hat{y} - \bar{y})^2}{\Sigma (y - \bar{y})^2}$$

Deze verhouding vertelt ons in welke mate de variatie van y verklaard wordt door de variatie van de regressielijn. Met andere woorden, we hebben een tweede grootheid om de sterkte van samenhang tussen twee variabelen te meten. Als r^2 groot is, zal de samenhang sterk zijn en in het omgekeerde geval zal de relatie zwak zijn.

In ons voorbeeld van de studenten vinden we nu voor r^2:

$r^2 = (0{,}88)^2 = 0{,}77$

0,77 betekent dat de variatie van de regressie 77 % is van de totale variatie of dat de variatie in de uitslagen voor wiskunde (X) 77 % van de variatie in de uitslagen voor informatica (Y) verklaart.

4 Hypothesetest voor ρ (rho)

Op basis van steekproefgegevens kunnen we nagaan of de correlatiecoëfficiënt van een populatie groter is dan nul. Om dat te illustreren, gebruiken we de steekproef correlatiecoëfficiënt r die we berekend hebben in het voorbeeld over de uitslagen voor wiskunde en informatica behaald door studenten.

We weten dat de correlatie tussen de uitslagen voor wiskunde en informatica positief zou moeten zijn, en dus formuleren we de volgende hypothese:

H_0: $\rho \leq 0$
H_1: $\rho > 0$
$\alpha = 0{,}05$

De formule voor de test is een t-test:

$$t = \frac{r - \rho}{s_r}$$

met $v = n - 2$

De noemer s_r is:

$$s_r = \sqrt{\frac{1 - r^2}{n - 2}}$$

De t-test ziet er uiteindelijk als volgt uit:

$$t = \frac{r - \rho}{\sqrt{\frac{1 - r^2}{n - 2}}} = \frac{r - \rho \times \sqrt{n - 2}}{\sqrt{1 - r^2}}$$

of voor het voorbeeld is t:

$$t = \frac{0{,}88 - 0 \times \sqrt{15 - 2}}{\sqrt{1 - 0{,}77}}$$

$$= \frac{0{,}88 \times 3{,}606}{0{,}48}$$

$$= 6{,}61$$

Voor de kritieke waarde van t bij v = 13 en α = 0,05 vinden we met de Excelfunctie T. INV een waarde van 1,77. Omdat t groter is dan t_k, besluiten we dat we H_0 verwerpen en dat ρ, de populatiecorrelatiecoëfficiënt, positief is.

5 Werkbladfuncties

Excel heeft een viertal werkbladfuncties die verband houden met correlatieberekeningen.

5.1 CORRELATIE (CORREL)

Figuur 109 toont de gegevens van het voorbeeld over de uitslagen voor wiskunde en informatica behaald door onze steekproefstudenten.

Hierop passen we de functie CORRELATIE toe:

> we selecteren cel B18 voor het resultaat van de Excelfunctie CORRELATIE;
> we klikken op de f_x-knop en selecteren in het venster 'Functie invoegen' de functie 'CORRELATIE';
> we klikken verder op 'OK' en het venster 'Functieargumenten' verschijnt;
> in het vak Matrix 1 voeren we het gegevensbereik B2:B16 in en in het vak Matrix 2 voeren we het bereik C2:C16 in;
> we klikken op 'OK' en het resultaat 0,88 verschijnt in cel B18.

B18 | =CORRELATIE(B2:B16;C2:C16)

	A	B	C	D	E
1	Student	Wiskunde(= X)	Informatica (= Y)		
2	1	70	73		
3	2	79	74		
4	3	85	81		
5	4	71	74		
6	5	88	82		
7	6	75	70		
8	7	90	82		
9	8	66	60		
10	9	86	90		
11	10	92	85		
12	11	74	77		
13	12	60	55		
14	13	69	72		
15	14	80	73		
16	15	77	72		
17					
18	CORRELATIE	0,88			
19					

Figuur 109 Berekening van r met CORRELATIE-functie

5.2 PEARSON (PEARSON)

Deze functie berekent de mate van lineaire samenhang tussen twee variabelen op dezelfde manier als de vorige functie. Om de PEARSON-functie te gebruiken, voeren we de volgende stappen uit:

> in Figuur 110 geven we de gegevens van ons voorbeeld opnieuw in en selecteren cel B18 voor het resultaat;
> we klikken op de f_x-knop, selecteren in het venster 'Functie invoegen' de functie 'PEARSON' en klikken op 'OK';
> in het venster 'Functieargumenten' voeren we in het vak Matrix 1 het bereik B2:B16 in en in het vak Matrix 2 het bereik C2:C16;
> we klikken op 'OK' en de correlatiecoëfficiënt r van Pearson verschijnt in cel B18 van het werkblad.

B18 =PEARSON(B2:B16;C2:C16)

	A	B	C	D	E
1	Student	Wiskunde(= X)	Informatica (= Y)		
2	1	70	73		
3	2	79	74		
4	3	85	81		
5	4	71	74		
6	5	88	82		
7	6	75	70		
8	7	90	82		
9	8	66	60		
10	9	86	90		
11	10	92	85		
12	11	74	77		
13	12	60	55		
14	13	69	72		
15	14	80	73		
16	15	77	72		
17					
18	Pearson	0,88			
19					

Figuur 110 Berekening van r met de PEARSON-functie

5.3 R. KWADRAAT (RSQ)

De functie R. KWADRAAT berekent de determinatiecoëfficiënt r^2.
Om deze functie toe te passen op het voorbeeld van de studenten, moet je de volgende stappen nemen:

> in Figuur 111 geef je in de bereiken B2:B16 en C2:C16 de gegevens in en selecteer je cel B18 voor de uitkomst;

> je klikt op de f_x-knop, selecteert in het venster 'Functie invoegen' de functie R. KWADRAAT en klikt op 'OK';
> in het venster 'Functieargumenten' breng je in het vak Y-bekend het bereik C2:C16 in;
> in het vak X-bekend voer je het bereik B2:B16 in;
> je klikt op 'OK' en in cel B18 verschijnt het antwoord voor r^2, zijnde 0,777.

B18 | =R.KWADRAAT(C2:C16;B2:B16)

	A	B	C	D	E
1	Student	Wiskunde(= X)	Informatica (= Y)		
2	1	70	73		
3	2	79	74		
4	3	85	81		
5	4	71	74		
6	5	88	82		
7	6	75	70		
8	7	90	82		
9	8	66	60		
10	9	86	90		
11	10	92	85		
12	11	74	77		
13	12	60	55		
14	13	69	72		
15	14	80	73		
16	15	77	72		
17					
18	R.KWADRAAT	0,777			
19					

Figuur 111 Berekening van de determinatiecoëfficiënt met de R. KWADRAAT-functie

5.4 COVARIANTIE. S (COVARIANCE. S)

De functie COVARIANTIE. S gebruikt de volgende formule:

$$\mathrm{cov}_{xy} = \frac{1}{n-1} \Sigma (x - \bar{x}) \cdot (y - \bar{y})$$

Hier volgen de stappen:

- in Figuur 112 voeren we eerst opnieuw de gegevens uit het voorbeeld in en kiezen cel B18 voor het resultaat van de functie;
- we klikken op de f_x-knop, selecteren in het venster 'Functie invoegen' de functie COVARIANTIE. S, klikken op 'OK' en het venster 'Functieargumenten' verschijnt;
- in het vak Matrix 1 voeren we het bereik B2:B16 in en in het vak Matrix 2 het bereik C2:C16;
- we klikken op 'OK' en in cel B18 verschijnt het resultaat, zijnde 74,88. Dit wijst op een positieve samenhang tussen beide variabelen.

B19 =COVARIANTIE.S(B2:B16;C2:C16)

	A	B	C	D	E	F
1	Student	Wiskunde (=X)	Informatica(Y)			
2	1	70	73			
3	2	79	74			
4	3	85	81			
5	4	71	74			
6	5	88	82			
7	6	75	70			
8	7	90	82			
9	8	66	60			
10	9	86	90			
11	10	92	85			
12	11	74	77			
13	12	60	55			
14	13	69	72			
15	14	80	73			
16	15	77	72			
17						
18						
19	COVARIANTIE.S	74,88				
20						

Figuur 112 Berekening van cov_{xy} met behulp van de functie COVARIANTIE. S

6 Gegevensanalysefunctie: Correlatie (Correlation)

Met deze functie kunnen we tegelijkertijd de berekening van meerdere correlaties uitvoeren als we te maken hebben met meer dan twee variabelen.

Als voorbeeld nemen we opnieuw de onderneming uit het vorige hoofdstuk. In Figuur 113 bekijken we een aantal gegevens over verkochte aantallen tv's (O = y), aantal advertenties (A = x_1) en toegepaste prijzen (P = x_2).

Volgende stappen voeren we uit om deze functie te gebruiken:

> we selecteren in het lint het tabblad 'Gegevens', daarna kiezen we 'Gegevensanalyse' en in het verschenen venster 'Gegevensanalyse' klikken we op de functie 'Correlatie'. We klikken op 'OK' en het venster 'Correlatie' wordt geopend;
> in het vak Invoerbereik voeren we het bereik A1:C11 in;
> we selecteren de optie 'Labels in de eerste rij', gezien de kolomtitels deel uitmaken van het bereik A1:C11;
> bij de uitvoeropties selecteren we 'Uitvoerbereik' en voeren cel A14 in om aan te duiden waar de uitvoer moet komen;
> we klikken op 'OK' en de uitvoertabel verschijnt in het bereik A14:D17.

	A	B	C	D	E
1	Advertering (A)	Prijs (P)	Omzet (O)		
2	3	1250	66		
3	6	1150	121		
4	10	1400	140		
5	13	1300	164		
6	9	1450	34		
7	6	1400	48		
8	4	1350	70		
9	5	1130	100		
10	9	1100	185		
11	4	1200	90		
12					
13					
14		*Advertering (A)*	*Prijs (P)*	*Omzet (O)*	
15	Advertering (A)	1			
16	Prijs (P)	0,251718978	1		
17	Omzet (O)	0,574060589	-0,523221201	1	
18					

Figuur 113 Uitvoer van de gegevensanalysefunctie Correlatie

De ingevulde cellen in de uitvoertabel geven telkens de correlatie tussen een variabele uit een rij en een variabele uit een kolom. In cel B17 bijvoorbeeld lezen we de sterkte van samenhang tussen de variabelen Advertenties en Omzet, die met een coëfficiënt van 0,57 vrij groot is.

We merken op dat de meervoudige correlatiecoëfficiënt R, of de coëfficiënt die de samenhang tussen een afhankelijke variabele en meerdere onafhankelijke variabelen beschrijft, al berekend werd met de gegevensanalysefunctie Regressie. In ons voorbeeld was R gelijk aan 0,897.

hoofdstuk 19

Trends

1 Voorspellingen

In dit hoofdstuk behandelen we drie technieken om uit een reeks gegevens een trend te ontwikkelen: het voortschrijdend gemiddelde, de exponentiële afvlakking en de methode van de kleinste kwadraten.

1.1 Voortschrijdend gemiddelde (Moving average)

Voor vele variabelen kan men in de tijd gegevens verzamelen. Die gegevens noemen we een tijdreeks.
Meestal gaan deze gegevens op en neer en is de fundamentele evolutie van de tijdreeksvariabele in kwestie niet duidelijk. Een mogelijke methode om de trend van de variabele duidelijk te maken, is het voortschrijdend gemiddelde.

Aan de hand van een voorbeeld leggen we deze techniek uit. Veronderstel dat we van een onderneming het jaarlijkse aantal personeelsleden van de laatste 15 jaar hebben bijgehouden. We willen nu deze reeks getallen vervangen door een reeks van voortschrijdende gemiddelden van n op elkaar volgende getallen. We stellen in ons voorbeeld n gelijk aan 3. Dat betekent dat als we de variabele 'aantal personeelsleden' gelijk stellen aan Y, de berekening van de voortschrijdende gemiddelden er als volgt uitziet:

$$\frac{y_1 + y_2 + y_3}{3} ; \frac{y_2 + y_3 + y_4}{3} ; \ldots ; \frac{y_{13} + y_{14} + y_{15}}{3}$$

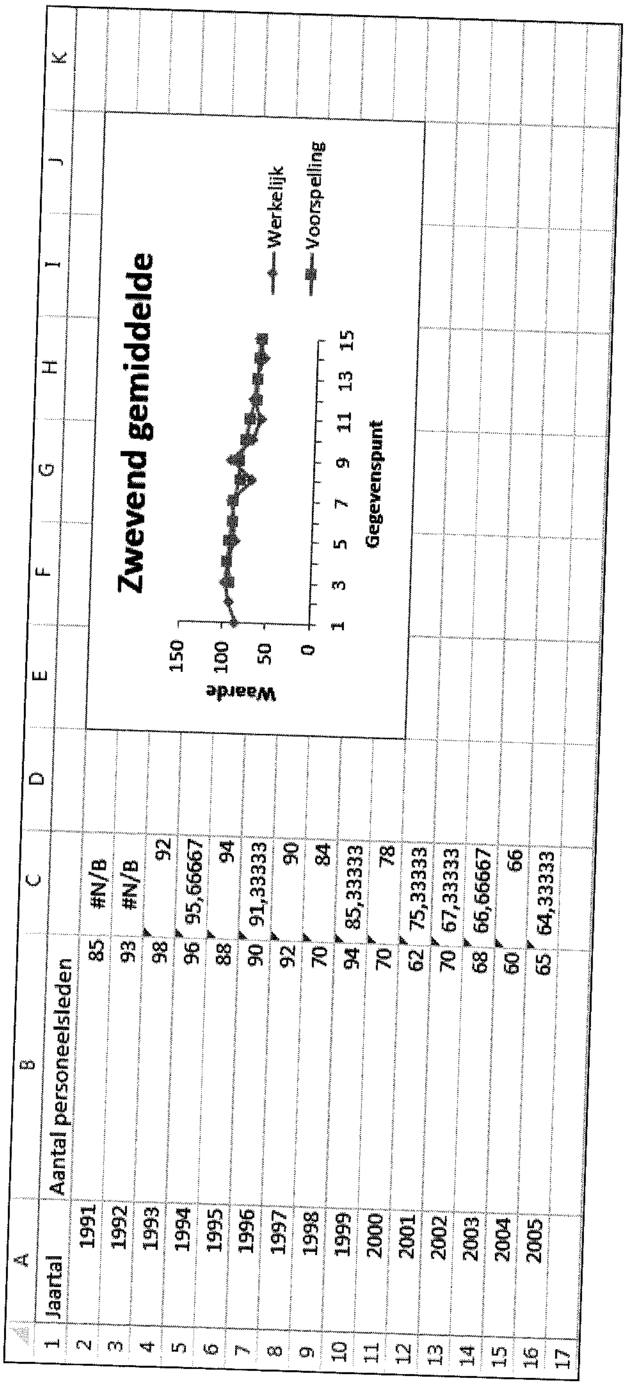
Excel beschikt over een gegevensanalysefunctie 'Zwevend gemiddelde', die de reeks van voortschrijdende gemiddelden berekent.

In Figuur 114 tonen we in het celbereik B2:B16 de personeelsaantallen van een onderneming voor de periode 1991 – 2005.
Hieronder tonen we het gebruik van de functie Zwevend gemiddelde (Moving average):

> we selecteren in het lint het tabblad 'Gegevens' en klikken daarna op de knop 'Gegevensanalyse';
> in het venster 'Gegevensanalyse' selecteren we de functie 'Zwevend gemiddelde';
> in het dialoogvenster 'Zwevend gemiddelde' vullen we in het vak Invoerbereik het gegevensbereik B1:B16 in;
> we selecteren de optie 'Labels in de eerste rij' omdat in het gegevensbereik de kolomtitel is opgenomen;
> in het vak Interval vullen we het getal 3 in, dat aangeeft hoeveel getallen we in het voortschrijdend gemiddelde willen opnemen;
> uit de opties 'Uitvoer' kiezen we de optie 'Uitvoerbereik' en voeren het bereik C2:C16 in voor het tonen van de voortschrijdende gemiddelden;
> een tweede Uitvoeroptie die we inschakelen is de optie 'Grafiek maken'.

Figuur 114 toont in kolom C de berekende voortschrijdende gemiddelden. Ieder voortschrijdend gemiddelde is een prognose op basis van de gegevens van de drie voorbije jaren. Bijvoorbeeld in cel C16 lezen we een voortschrijdend gemiddelde van 64,33 af, dat betekent dat op basis van de laaste drie jaren het verwachte aantal personeelsleden voor het jaar 2005 gelijk is aan 64,33.

De grafiek van de trendwaarden toont duidelijk dat de schommelingen in de werkelijke Y-waarden worden afgevlakt, doordat de berekening van jaarlijkse trendwaarden de toevallige schommelingen elimineert.



	A	B	C
1	Jaartal	Aantal personeelsleden	
2	1991	85	#N/B
3	1992	93	#N/B
4	1993	98	92
5	1994	96	95,66667
6	1995	88	94
7	1996	90	91,33333
8	1997	92	90
9	1998	70	84
10	1999	94	85,33333
11	2000	70	78
12	2001	62	75,33333
13	2002	70	67,33333
14	2003	68	66,66667
15	2004	60	66
16	2005	65	64,33333
17			

Figuur 114 De voortschrijdende gemiddelden en hun grafiek met de gegevensanalysefunctie Zwevend gemiddelde

1.2 Exponentiële afvlakking (Exponential smoothing)

Exponentiële afvlakking is een voorspellingstechniek die een trendwaarde schat op basis van de trendwaarde en de werkelijke waarde van de vorige periode.

De formule voor T_{t+1}:

$$T_{t+1} = \alpha\, T_t + (1 - \alpha)\, Y_t$$

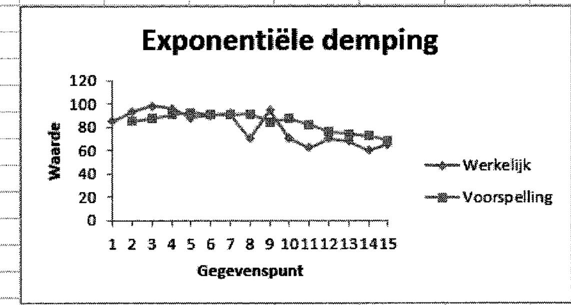
Toegepast op ons voorbeeld van de personeelsaantallen is T_{t+1} het geschatte aantal personeelsleden voor t + 1. Y_t is het werkelijke aantal personeelsleden in periode t, T_t is het geschatte aantal personeelsleden voor periode t en α is de afvlakkingsconstante waarvan de grootte bepaalt in welke mate er rekening wordt gehouden met T_t. Hoe groter α, hoe meer er rekening wordt gehouden met de voorgaande voorspelling.

De gegevensanalysefunctie ***Exponentiële demping*** (Exponential smoothing) laat ons toe de trendwaarden en de grafiek ervan voor ons voorbeeld snel te produceren. Daarbij ga je als volgt te werk:

> in het werkblad van Figuur 115 hebben we eerst de gegevens hernomen van het voorbeeld;
> in het lint selecteren we het tabblad 'Gegevens', klikken dan op 'Gegevensanalyse' en schakelen de functie 'Exponentiële demping' uit het dialoogvenster 'Gegevensanalyse' in;
> het venster 'Exponentiële demping' verschijnt en in het vak 'Invoerbereik' brengen we het gegevensbereik B2:B16 in;
> daarna vullen we in het vak Dempingsfactor een α-waarde in gelijk aan 0,70;
> in het Uitvoerbereik voeren we het bereik C2:C16 in en als bijkomende uitvoeroptie selecteren we ook nog 'Grafiek maken';
> we klikken op 'OK' en het resultaat verschijnt in het werkblad.

We stellen vast dat met een afvlakkingsconstante gelijk aan 0,7 de instabiliteit in de Y-waarden in belangrijke mate wordt weggenomen.

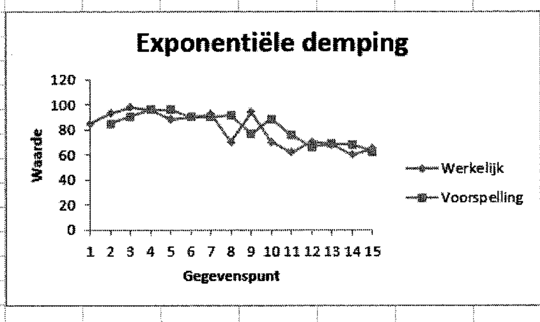
	A	B	C
1	Jaartal	Aantal personeelsleden	
2	1991	85	#N/B
3	1992	93	85
4	1993	98	87,4
5	1994	96	90,58
6	1995	88	92,206
7	1996	90	90,9442
8	1997	92	90,66094
9	1998	70	91,06266
10	1999	94	84,74386
11	2000	70	87,5207
12	2001	62	82,26449
13	2002	70	76,18514
14	2003	68	74,3296
15	2004	60	72,43072
16	2005	65	68,7015
17			



Figuur 115 Berekening van trendwaarden met de gegevensanalysefunctie Exponentiële demping en α = 0,70

Maar als we uitgaan van α = 0,30, zijnde een gebruikelijke standaardwaarde voor α, dan zien we in ons voorbeeld dat de trendgrafiek in Figuur 116 een grotere gelijkenis vertoont met de grafiek van de werkelijke Y-waarden. Dat is het gevolg van het toekennen van een groter gewicht aan de werkelijke Y-waarden.

	A	B	C
1	Jaartal	Aantal personeelsleden	
2	1991	85	#N/B
3	1992	93	85
4	1993	98	90,6
5	1994	96	95,78
6	1995	88	95,934
7	1996	90	90,3802
8	1997	92	90,11406
9	1998	70	91,43422
10	1999	94	76,43027
11	2000	70	88,72908
12	2001	62	75,61872
13	2002	70	66,08562
14	2003	68	68,82569
15	2004	60	68,24771
16	2005	65	62,47431
17			



Figuur 116 Berekening van trendwaarden met de gegevensanalysefunctie Exponentiële demping en α = 0,30

1.3 Methode van de kleinste kwadraten

1.3.1 Definitie

In een onderneming kan men gedurende een bepaalde periode, bijvoorbeeld op het einde van ieder kwartaal, de omzet van een product noteren. Deze, op vaste tijdstippen genoteerde omzetgegevens, genereren een reeks getallen die we een tijdreeks noemen. Typisch aan zo'n tijdreeks is dat de tijdsintervallen tussen de opeenvolgende gegevens gelijk zijn.

1.3.2 Schatten van de trend

We willen de evolutie van een tijdreeks over een langere termijn, of met andere woorden, de trend van een variabele in functie van de tijd, grafisch voorstellen. Is de trendlijn een rechte lijn dan kun je de vergelijking ervan vinden door toepassing van de methode van de kleinste kwadraten.
Dit is een techniek die ons toelaat om een richtingscoëfficiënt a en een constante b te berekenen zodat de werkelijke punten van de variabele y minimaal afwijken van de punten op de trendlijn.

De vergelijking van een rechte is:

$y = ax + b$

Gezien x hier gelijk is aan t of de tijd, en y gelijk is aan de trendwaarde T, wordt de vergelijking:

$T = at + b$

De waarden van a en b worden gevonden door volgende formules:

$$a = \frac{\sum(t-\bar{t})(y-\bar{y})}{\sum(t-\bar{t})^2}$$

$$b = \bar{y} - a\bar{t}$$

waarbij:

y = werkelijke waarde van de waarneming
t = gecodeerde tijdseenheid

In Figuur 117 tonen we de evolutie van de kwartaalomzetcijfers van een onderneming gerealiseerd in de periode 20x1-20x5:

	A	B	C
1	**Kwartaal**	**y**	
2	20x1-1	75	
3	20x1-2	80	
4	20x1-3	90	
5	20x1-4	85	
6	20x2-1	95	
7	20x2-2	90	
8	20x2-3	105	
9	20x2-4	100	
10	20x3-1	110	
11	20x3-2	100	
12	20x3-3	120	
13	20x3-4	115	
14	20x4-1	125	
15	20x4-2	110	
16	20x4-3	140	
17	20x4-4	130	
18	20x5-1	145	
19	20x5-2	140	
20	20x5-3	160	
21	20x5-4	150	
22			

Figuur 117 Evolutie kwartaalomzetcijfers 20x1-20x5 (000 euro)

Wat betreft de codering van t laten we het nulpunt op de t-as steeds samenvallen met het midden van de te beschouwen periode, hierdoor is $\bar{t}$ steeds gelijk aan nul en wordt $b = \bar{y}$.
In het geval dat het aantal tijdseenheden even is, moeten we natuurlijk het nulpunt tussen de twee middelste tijdseenheden leggen.

In Figuur 118 berekenen we a en b en de vergelijking van T.

	A	B	C	D	E	F	G	H	I
1									
2		Kwartaal	t	y	$t-\bar{t}$	$y-\bar{y}$	$(t-\bar{t})\cdot(y-\bar{y})$	$(t-\bar{t})^2$	
3		20x1-1	-19	75	-19	-38,25	726,75	361	
4		20x1-2	-17	80	-17	-33,25	565,25	289	
5		20x1-3	-15	90	-15	-23,25	348,75	225	
6		20x1-4	-13	85	-13	-28,25	367,25	169	
7		20x2-1	-11	95	-11	-18,25	200,75	121	
8		20x2-2	-9	90	-9	-23,25	209,25	81	
9		20x2-3	-7	105	-7	-8,25	57,75	49	
10		20x2-4	-5	100	-5	-13,25	66,25	25	
11		20x3-1	-3	110	-3	-3,25	9,75	9	
12		20x3-2	-1	100	-1	-13,25	13,25	1	
13		20x3-3	1	120	1	6,75	6,75	1	
14		20x3-4	3	115	3	1,75	5,25	9	
15		20x4-1	5	125	5	11,75	58,75	25	
16		20x4-2	7	110	7	-3,25	-22,75	49	
17		20x4-3	9	140	9	26,75	240,75	81	
18		20x4-4	11	130	11	16,75	184,25	121	
19		20x5-1	13	145	13	31,75	412,75	169	
20		20x5-2	15	140	15	26,75	401,25	225	
21		20x5-3	17	160	17	46,75	794,75	289	
22		20x5-4	19	150	19	36,75	698,25	361	
23							5345		
24									
25									
26			$\bar{y}$	113,25					
27			$\bar{t}$	0					
28									
29			$\Sigma(t-\bar{t})\cdot(y-\bar{y})$		5345				
30									
31			$\Sigma\,(t-\bar{t})^2$		2660				
32									
33			a		2,009398				
34									
35			b		113,25				
36									
37			T = 2,0094 t + 113,25						
38									

Figuur 118 Berekening van a, b en de trendvergelijking T

Met de vergelijking van T berekenen we in Figuur 119 de trendwaarden voor de kwartalen van 20X1-1 tot en met 20X5-4.

	A	B	C	D
1	t	Y	T	
2	-19	75	75,07	
3	-17	80	79,09	
4	-15	90	83,11	
5	-13	85	87,13	
6	-11	95	91,15	
7	-9	90	95,17	
8	-7	105	99,18	
9	-5	100	103,20	
10	-3	110	107,22	
11	-1	100	111,24	
12	1	120	115,26	
13	3	115	119,28	
14	5	125	123,30	
15	7	110	127,32	
16	9	140	131,33	
17	11	130	135,35	
18	13	145	139,37	
19	15	140	143,39	
20	17	160	147,41	
21	19	150	151,43	
22				

Figuur 119 Berekening van T-waarden (000 euro) met T= 2,0094t + 113,25

Met de trendvergelijking kunnen we voor latere kwartalen ook voorspellingen berekenen:

	A	B	C	D
1	Kwartaal	t	T	
2	20X6-1	21	155,45	
3	20X6-2	23	159,47	
4	20X6-3	25	163,49	
5	20X6-4	27	167,50	
6				

Figuur 120 Berekening van trendwaarden (000 euro) voor de kwartalen van 20X6

1.3.3 Trendlijn en trendvergelijking met Excel

We stellen een tabel op met de omzetcijfers die in de gecodeerde kwartalen van 20X1 tot en met 20X5 door een onderneming werden gerealiseerd:

	A	B	C
1	t	y	
2	-19	75	
3	-17	80	
4	-15	90	
5	-13	85	
6	-11	95	
7	-9	90	
8	-7	105	
9	-5	100	
10	-3	110	
11	-1	100	
12	1	120	
13	3	115	
14	5	125	
15	7	110	
16	9	140	
17	11	130	
18	13	145	
19	15	140	
20	17	160	
21	19	150	
22			

Figuur 121 Omzetgegevens in duizenden euro's per kwartaal voor de periode 20X1-20X5

Om de trendlijn en de trendvergelijking met Excel te bepalen voer je de volgende stappen uit:

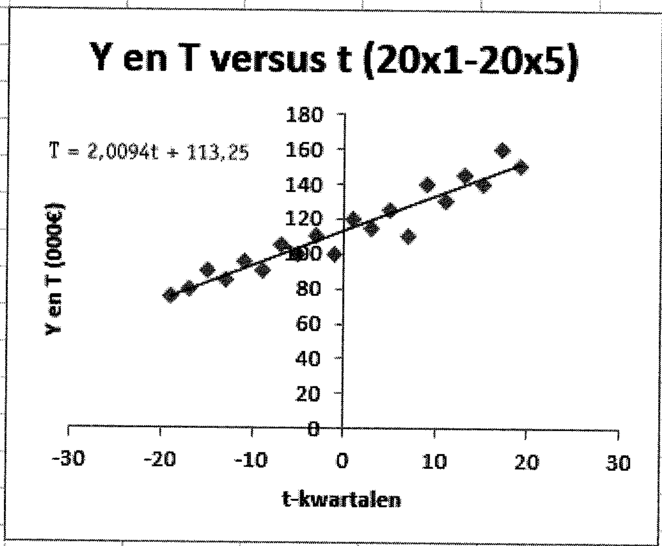
> selecteer het gegevensbereik A2:B21, klik op het tabblad 'Invoegen', kies in de groep 'Grafieken' de keuzelijst 'Spreidings- (X, Y) of bellendiagram invoegen' en selecteer het type 'Spreiding'. Het resultaat is een spreidingsdiagram van een tijdreeks van Y-waarden;

> klik vervolgens met de rechtermuisknop op een punt in het spreidingsdiagram. In de lijst die verschijnt selecteer je 'Trendlijn toevoegen'. Klik in het dialoogvenster 'Trendlijn opmaken' op de optie 'Lineair' en vink ook de optie 'Vergelijking in grafiek weergeven' aan. In het diagram verschijnt een trendlijn en een lineaire vergelijking. Vervang y en x door respectievelijk T en t in de vergelijking;

> klik dan in de groep 'Grafiekindelingen' op 'Grafiekonderdeel toevoegen' en selecteer 'Astitels'. Klik op de optie 'Primair horizontaal' en voer in het vak 'Astitel' de titel 't-kwartalen' in. Selecteer 'Astitels' opnieuw, klik op 'Primair verticaal' en voer in het vak 'Astitel', 'Y en T (000 €)' in. Selecteer in de keuzelijst 'Grafiekonderdeel toevoegen' vervolgens nog het onderdeel 'Grafiektitel' en kies de optie 'Boven grafiek'. In het label 'Grafiektitel' vul je 'Y en T versus t (20X1 - 20X5)' in;
> verwijder ten slotte de rasterlijnen. Selecteer in de keuzelijst 'Grafiekonderdeel toevoegen' het onderdeel 'Rasterlijnen' en klik op de optie 'Primair groot horizontaal' en vervolgens op de optie 'Primair groot verticaal'.

In Figuur 122 zien we het resultaat.

	A	B
1	t	y
2	-19	75
3	-17	80
4	-15	90
5	-13	85
6	-11	95
7	-9	90
8	-7	105
9	-5	100
10	-3	110
11	-1	100
12	1	120
13	3	115
14	5	125
15	7	110
16	9	140
17	11	130
18	13	145
19	15	140
20	17	160
21	19	150
22		



Figuur 122 Spreidingsdiagram, trendlijn en trendvergelijking

Voor T-waarden van de kwartalen na 20X5 moeten we een gecodeerde t-waarde in de vergelijking invullen. Voor het kwartaal 20X6-1 vinden we dan bijvoorbeeld (t = 21):

$$T = 2{,}009 \times 21 + 113{,}25 = 155{,}44$$

Bijlagen

Bijlage 1 Oefeningen

Hoofdstuk 3 & 4

Oefening 1

Hierna vind je de uitslagen van 20 studenten (op 20 punten) voor het vak informatica:

16	11	17	10	17
15	9	15	14	15
12	16	6	12	8
10	3	9	7	19

De intervallen zijn 4, 8, 12, 16 en 20.

Gevraagd:
> stel de frequentieverdeling op;
> teken een kolommendiagram;
> bereken het gemiddelde, de mediaan en de modus.

Oefening 2

De onderstaande tabel geeft een trimestrieel overzicht van de productiehoeveelheden van de verschillende productgroepen van een onderneming voor het jaar 20X1.

Tabel Trimestriële productie-evolutie van de diverse productgroepen in het jaar 20X1 (000 eenheden)

Productgroep	Trim 1	Trim 2	Trim 3	Trim 4
Fietsen	58	65	60	52
Racefietsen	10	11	13	8
Mountainbikes	4	7	9	3
Bromfietsen	6	8	7	6

Gevraagd:

> maak een kolommendiagram van de trimestriële productie van de productgroepen in 20X1;
> vul de bovenstaande tabel aan met de totale producties per trimester en maak een staafdiagram van de trimestriële productiehoeveelheden in 20X1;
> maak een cirkelgrafiek met de procentuele aandelen van de verschillende productgroepen in de productie van het tweede trimester van 20X1;
> teken een lijngrafiek die de trimestriële evolutie weergeeft van de productie van fietsen in 20X1.

Oefening 3

De volgende tabel geeft de leeftijden X en de systolische bloeddrukken Y van 10 mensen:

Leeftijd X	66	32	74	55	60	39	40
Bloeddruk Y	170	118	165	150	152	121	137
Leeftijd X	28	70	49				
Bloeddruk Y	110	154	146				

Gevraagd:

> teken het spreidingsdiagram;
> teken de trend- of regressielijn;
> bepaal de regressievergelijking.

Oefening 4

De investeringsmaatschappij Robelco deelt in haar verslag van 20X6 mee dat het rendement over de laatste vijf jaar gemiddeld 8,53 % was.
Deze vaststelling is juist als de rendementen in 20X1, 20X2, 20X3, 20X4 en 20X5:

a	7 %	10 %	8 %	5 %	9 % waren;
b	3 %	15 %	10 %	9 %	6 % waren;
c	14 %	5 %	7 %	10 %	12 % waren?

Oefening 5

In 5 opeenvolgende jaren geeft een onderneming jaarlijks € 5 000 uit aan transportkosten. De kosten per kilometer bedroegen respectievelijk € 0,10, € 0,15, € 0,20, € 0,22 en € 0,25. Hoe groot is het harmonisch gemiddelde van de transportkost per kilometer?

Hoofdstuk 5

Oefening 6

De directeur van een klein hotel heeft in de onderstaande tabel het aantal dagen verblijf in zijn hotel geregistreerd van alle klanten tijdens de zomerperiode van 20X1:

5	15	7	4	11	23
4	5	12	1	6	20
3	8	20	9	5	11
6	6	11	10	2	21
10	9	6	17	2	19

De intervallen zijn 5, 10, 15, 20 en 25.

Gevraagd:
> stel de frequentieverdeling op;
> bereken het gemiddelde;
> bereken de populatievariantie en de populatiestandaarddeviatie;
> teken een kolommendiagram van de frequentieverdeling.

Oefening 7

Een steekproef van 40 batterijen levert de volgende gebruikstijden op (in uren):

50	58	50	51	51
53	59	13	35	62
47	41	46	40	45
52	49	49	57	56
50	62	53	49	50
17	46	52	60	56
55	53	51	57	64
12	58	57	61	29

Gevraagd:
> bereken het gemiddelde;
> bereken het getrimd gemiddelde als we 20 % van de observaties als extreem beschouwen;
> bereken de steekproefvariantie en steekproefstandaardafwijking;
> stel de frequentieverdeling op. De intervallen zijn 10, 20, 30, 40, 50, 60 en 70.

Hoofdstuk 6

Oefening 8

Standaardiseer de uitslagen van Oefening 1 en beschouw ze als een populatie.

Oefening 9

Een student behaalde 64 op een examen in economie waarvoor de gemiddelde score 70 was en de standaarddeviatie 8. Op het examen van statistiek waarvoor de gemiddelde score 57 was en de standaarddeviatie gelijk was aan 12, behaalde hij 60.

Voor welk vak is de relatieve positie van de student het best?

Oefening 10

Bereken voor de uitslagen opgegeven in Oefening 1:
> de rang in stijgende volgorde;
> de 6de grootste uitslag en de 4de kleinste uitslag;
> het 60ste, 50ste en 95ste percentiel;
> de relatieve positie van de uitslagen 10 en 12.

Oefening 11

Hieronder vind je de lichaamsgewichten (kg) van de leden van een sportclub:

65	80	74	68	85
70	72	66	62	63
60	69	83	71	77
59	50	55	79	61

Gevraagd:
Bereken alle kwartielen.

Oefening 12

Bereken met de gegevensanalysetechniek Rang en percentiel van de gebruikstijden van de 40 batterijen opgegeven in Oefening 7 hun rang en percentiel.

Hoofdstuk 7

Oefening 13

De gemiddelde prijs van een grondstof per kg is in onderstaande tabel voor de jaren 20x0-20x5 gegeven.

Gevraagd:
> gebruik 20x0 als basisjaar en bereken de prijsindexcijfers voor de jaren 20x3 en 20x5;
> gebruik 20x3 als basisjaar en bereken de prijsindexcijfers voor alle jaren.

Jaar	20x0	20x1	20x2	20x3	20x4	20x5
Gemiddelde prijs (€)	10,50	10,60	10,65	10,75	10,92	11,00

Oefening 14

In de volgende tabel vind je de prijzen en productie van drie voedingsproducten voor de jaren 20x0, 20x1 en 20x2. Bereken een samengesteld prijsindexcijfer voor het jaar 20x2 waarbij je het jaar 20x0 als basisjaar gebruikt. Maak dezelfde berekening waarbij je 20x0-20x1 als basisjaar gebruikt.

Product	P 20x0	P 20x1	P 20x2	Q 20x0	Q 20x1	Q 20x2
A	€ 4,00	€ 3,80	€ 4,10	9 000 units	9 250 units	9 500 units
B	€ 10,20	€ 12,50	€ 9,00	6 000 units	5 750 units	5 750 units
C	€ 52,00	€ 55,00	€ 60	7 000 units	6 500 units	7 100 units

Oefening 15

Gebruik de gegevens van oefening 14 en bereken een Laspeyres-prijsindexcijfer voor het jaar 20x2 met 20x0 als basisjaar.

Oefening 16

Bereken het Fisher-prijsindexcijfer voor de drie voedingsproducten van oefening 14 voor het jaar 20x2 met als basisjaar 20x0.

Oefening 17

De volgende tabel geeft de prijzen en hoeveelheden weer voor salons en eetkamers geproduceerd door nv Meubelparadijs tijdens de jaren 20x0 en 20x9.

Product	Prijs 20x0	Prijs 20x9	Hoeveelheden 20x0	Hoeveelheden 20x9
Salons	€ 2 500	€ 3 750	25 000	15 000
Eetkamers	€ 3 250	€ 4 000	14 000	17 500

Bereken een hoeveelheidsindexcijfer voor 20x9 met als basisjaar 20x0 door middel van:

a een gewogen hoeveelheidsindexcijfer met de prijzen van het basisjaar als gewichten.

b een gewogen hoeveelheidsindexcijfer met de prijzen van het laatste jaar als gewichten.

Oefening 18

Bereken het waarde-indexcijfer voor de gegevens uit oefening 17.

Oefening 19

De volgende prijzen zijn gegeven voor een elektrische fiets verkocht door bvba Easy Cycling in de periode 20x0-20x4:

Jaar	Prijs (euro's)
20x0	1 500
20x1	1 300
20x2	1 200
20x3	1 000
20x4	890

Gevraagd:

> Geef de schakelverhoudingen;

> Bereken de prijsindexcijfers als het basisjaar 20x2 is.

Hoofdstuk 8

Oefening 20

We kennen de volgende reeks van gegevens:

6	14	78
16	28	66
25		55
	36	40
34	44	27
19	ja	
25	neen	30
Ja		25
neen	37	ja

Bereken:

> het aantal cellen dat getallen bevat;
> het aantal cellen dat getallen en tekst bevat;
> het aantal lege cellen;
> het aantal cellen dat voldoet aan het criterium ≥ 19;
> het grootste en het kleinste getal.

Hoofdstuk 9

Oefening 21

Hieronder vind je de jaarlijkse inkomens (in 000 euro) van 40 bedienden van de onderneming Comex nv:

20	22	15	55	57
25	36	23	50	50
27	38	37	51	28
30	41	48	38	33
41	52	46	37	19
50	36	44	36	42
38	34	50	40	18
23	54	51	44	48

Gevraagd:

> stel een frequentieverdeling op met als intervallen: 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60;
> teken een histogram;
> bereken de scheefheid van de frequentieverdeling.

Oefening 22

De volgende gegevens geven een overzicht van de afgelegde kilometers het voorbije jaar door de handelsreizigers die een bedrijfswagen van de onderneming mochten gebruiken:

7 200	4 200	16 000	7 300	13 000	7 300
14 000	8 100	19 800	11 800	20 600	9 250
19 000	9 500	12 100	26 200	29 000	11 800
22 000	13 800	10 500	24 000	7 000	16 500
27 000	4 700	8 900	4 000	15 500	22 600

Gevraagd:

> maak met de gegevensanalysefunctie Histogram een frequentieverdeling, een kolommendiagram en een cumulatieve relatieve frequentiegrafiek. De intervallen zijn 5 000, 10 000, 15 000, 20 000, 25 000 en 30 000;

> beschrijf de frequentieverdeling met de functie Beschrijvende statistiek.

Hoofdstuk 10

Oefening 23

Op hoeveel manieren kunnen 10 verschillende gekleurde brieven geordend worden?

Oefening 24

Op hoeveel verschillende manieren kunnen 7 treinreizigers op een zitbank plaatsnemen als er slechts 3 zitplaatsen beschikbaar zijn?

Oefening 25

Vijf wielertoeristen moeten aan hun eigen fiets toegewezen worden. Er is maar één juiste toewijzing. De fietsers worden lukraak toegewezen.

Gevraagd:

Bereken de kans dat er 5, 1, 2 fietsers juist worden toegewezen aan hun fiets.

Oefening 26

Op een schap staan 15 boeken waarvan er 9 gaan over statistiek en 6 over financiële wiskunde. Wat is de kans dat als men 4 boeken willekeurig van het schap neemt:

a de 4 genomen boeken over statistiek gaan;
b er 1 boek handelt over statistiek en de 3 andere over financiële wiskunde gaan;
c de 4 genomen boeken over financiële wiskunde gaan?

Oefening 27

Op hoeveel verschillende manieren kun je een jury van 5 mensen selecteren uit 10 mensen?

Oefening 28

Uit een groep van 5 bedrijfsmensen en 10 docenten moet een examencommissie worden samengesteld, van 3 bedrijfsmensen en 4 docenten. Op hoeveel verschillende manieren kan dat gebeuren als:

a alle bedrijfsmensen en docenten geselecteerd kunnen worden;
b twee specifieke bedrijfsmensen geen deel kunnen uitmaken van de commissie?

Oefening 29

Een dobbelsteen wordt 2 keer opgeworpen. Bepaal de kans om een 1 of een 3 te werpen bij een eerste worp en een 4 of een 6 bij een tweede worp.

Oefening 30

Er worden twee kaarten uit een gewoon kaartspel getrokken. Wat is de kans dat de trekker:

a 2 tienen trekt?
b 2 zwarte tienen trekt?
c 1 heer en 1 dame trekt?

Oefening 31

Drie muntstukken worden opgeworpen en we bepalen dat:

A = de eerste twee muntstukken leveren munt op
B = het laatste muntstuk levert munt op
C = alle muntstukken leveren munt op

Gevraagd:

> Zijn A en B onafhankelijk?
> Zijn A en C onafhankelijk?

Oefening 32

Drie renners, X, Y en Z, nemen deel aan een wedstrijd op de piste. De kansen van X en Y worden gelijk geacht en zij hebben dubbel zo weinig kans als Z.

a Bereken de kans P(X), P(Y) en P(Z).
b Bereken de kans dat X of Z wint.

Hoofdstuk 11

Oefening 33

Mevrouw Verkest heeft een textielzaak. Ze heeft berekend dat de kans dat een klant die haar winkel bezoekt iets koopt, gelijk is aan 0,25.

Als we aannemen dat er om het uur 7 klanten haar winkel bezoeken:

> stel de binomiale kansverdelingstabel op;
> teken de grafiek van deze binomiale kansverdeling;
> wat is de kans dat er hoogstens 3 klanten iets kopen in een bepaald uur;
> wat is de kans dat er 5 of 6 klanten iets kopen in een bepaald uur?

Oefening 34

Bereken de kans dat een student die deelneemt aan een examen met twintig juist-fout-vragen, van die twintig gestelde vragen:

> er 10 juist beantwoordt;
> er minstens 18 juist beantwoordt;
> er hoogstens 10 juist beantwoordt.

Oefening 35

De beste van twee basketbalploegen heeft 0,65 kans om in iedere wedstrijd tegen de andere ploeg te winnen. Als er 6 keer gespeeld wordt, wat is dan de kans dat de beste ploeg hoogstens 3 van de 6 wedstrijden zal winnen?

Oefening 36

Een dobbelsteen wordt 25 keer opgeworpen en levert 7 keer 6 ogen op. Kunnen we besluiten bij een foutenrisico $\alpha = 0{,}05$ en $\alpha = 0{,}02$ dat de dobbelsteen onzuiver is?

Oefening 37

Vijf procent van de afgewerkte goederen wordt op het einde van een productieproces gewoonlijk afgekeurd.

Bereken de kans dat in een steekproef van 100 eindproducten:
> er juist 15 worden afgekeurd;
> er hoogstens 5 worden afgekeurd.

Oefening 38

Een producent van koelkasten levert een partij aan een klant. Gemiddeld is 2 % defect. Beide partijen komen overeen dat er een steekproef van 10 eenheden wordt genomen. Als er 3 of minder apparaten worden afgekeurd, wordt de levering aanvaard.

Bereken de kans dat de koelkasten aanvaard zullen worden.

Hoofdstuk 12

Oefening 39

Een chocoladefabriek produceert chocoladerepen van 100 g met een gemiddeld chocoladegehalte van 75 g en een standaarddeviatie van 0,95 g.

Wat is de kans dat een willekeurig geselecteerde chocoladereep:
> meer dan 78 g chocolade bevat;
> minder dan 73 g chocolade bevat;
> tussen 74 en 77 g chocolade bevat?

Oefening 40

De gemiddelde score op een eindexamen was 70 en de standaarddeviatie 10.
De top 5 % van de studenten ontvangt een waardebon van € 500.

Hoe groot is de minimumscore die een student moet hebben om een waardebon te kunnen krijgen?

Oefening 41

De gewichtsverliezen van de leden van een vermageringsclub zijn normaal verdeeld met een gemiddelde van 8,75 kg en een standaarddeviatie van 1,82 kg.

> Welk percentage van de leden mag verwachten dat de gewichtsafname meer dan 9 kg zal bedragen?
> Welk percentage van de leden mag verwachten dat de gewichtsafname minder dan 5 kg zal bedragen?
> Wat is het hoogste gewicht dat de 40 % minst succesvolle afslankers mogen verwachten te verliezen?
> Wat is het laagste gewicht dat de 30 % meest succesvolle afslankers mogen verwachten te verliezen?

Oefening 42

Als μ het gemiddelde is en σ de standaardafwijking van een populatie die normaal verdeeld is, welk percentage van de waarnemingen ligt:

> binnen het interval $\mu \pm 2\sigma$;
> buiten het interval $\mu \pm 2{,}5\sigma$?

Hoofdstuk 13

Oefening 43

Een steekproef van 34 dozen melk wordt genomen uit een schap van 1 000 dozen in het magazijn van een zuivelfabriek. De vermelde inhoud is 1 l.
De waargenomen inhouden zijn:

Gewicht in l	0,9	1,05	1,10	1,15
Aantal dozen	10	16	6	2

Bereken het 95 %-betrouwbaarheidsinterval voor de gemiddelde inhoud van alle dozen op het schap.

Oefening 44

De bestuursraad van een middelbare school beschouwt het als haar belangrijkste taak het gemiddelde aantal leerlingen per klas lager te houden dan het gemiddelde aantal leerlingen in een naburige school.

De schooldirecteur deelt aan de bestuursraad mee dat het gemiddelde aantal leerlingen per klas bij de concurrerende school 31,4 leerlingen bedraagt. In de eigen school zijn voor 40 van de 100 klassen de leerlingen geteld. Het klasgemiddelde bedraagt 28. Verder is s gelijk aan 7 leerlingen.

> Bereken het interval waarbij de schooldirecteur 97 % zekerheid heeft dat het populatiegemiddelde erin begrepen is;
> Denk je dat de schooldirecteur/bestuursraad zijn/haar doelstelling heeft gerealiseerd?

Oefening 45

De heer Janssens is een beursmakelaar op Euronext en wil weten hoeveel tijd er verloopt tussen het plaatsen en uitvoeren van een beursorder. Hij stelde op basis van 45 uitgevoerde orders vast dat de gemiddelde uitvoeringstijd 25 minuten is en dat de standaarddeviatie 3,5 minuten bedraagt.

Bereken het 98 %-betrouwbaarheidsinterval van de gemiddelde uitvoeringstijd.

Oefening 46

De politiecommissaris stelt een onderzoek in naar het grote aantal speelpleinongevallen die tijdens de recente zomers plaatsvonden in zijn gemeente. Hij selecteerde willekeurig 9 zomermaanden uit de laatste jaren en verzamelde gegevens over het aantal speelpleinongevallen in de geobserveerde maanden.
Het gemiddelde aantal speelpleinongevallen in die 9 maanden was 31 en de standaardafwijking bedraagt 9 ongevallen per maand.

Bereken voor de commissaris een 90 %-betrouwbaarheidsinterval voor het populatiegemiddelde van speelpleinongevallen per maand.

Oefening 47

Men test de reactiesnelheid van een persoon op een bepaalde prikkel. Men doet 5 metingen met als uitkomsten: 0,28; 0,30; 0,27; 0,33 en 0,31 seconden.

Bereken de 99 %-zekerheidsgrenzen van de werkelijke reactietijd van deze persoon.

Hoofdstuk 14

Oefening 48

De nv Martin voert een speciale handelspromotie voor de verkoop van houtkachels en is van oordeel dat de promotie moet resulteren in een prijsverandering voor de consument. Vóór de promotie begon, was de gemiddelde prijs van een kachel € 449,5 en de standaarddeviatie € 57,5.
De onderneming deed onderzoek bij 25 handelaars nadat de promotie was begonnen en constateerde een gemiddelde prijs van € 429,5.

a Met een foutenrisico van $\alpha = 0{,}02$ kan de nv Martin ervan uitgaan dat de gemiddelde kleinhandelsprijs voor de consument gedaald is?
b Met een foutenrisico van $\alpha = 0{,}02$ kan de nv Martin ervan uitgaan dat de gemiddelde kleinhandelsprijs voor de consument veranderd is?

Oefening 49

Het gemiddelde commissieloon aangerekend door beursvennootschappen op de verkoop van aandelen is € 144 en de standaarddeviatie is € 52. Jan Verbist nam een willekeurige steekproef van 121 transacties bij zijn klanten en stelde vast dat zij een gemiddeld commissieloon van € 151 betaalden.

Kan Jan Verbist, met een foutenrisico van $\alpha = 0{,}10$, besluiten dat de commissies aangerekend aan zijn klanten hoger zijn dan het sectorgemiddelde?

Oefening 50

Er werd een examen gegeven aan studenten van verschillende scholen. De gemiddelde uitslag was 74,5 en de standaarddeviatie was 8. In een bepaalde school waar 200 studenten deelnamen aan het examen, was de gemiddelde uitslag 75,9.

Met een foutenrisico van $\alpha = 0{,}05$ kunnen we besluiten:
a dat het gemiddelde resultaat behaald in deze school beter is dan de resultaten behaald in de andere scholen;
b dat het gemiddelde resultaat behaald in deze school verschillend is van de resultaten behaald in de andere scholen?

Oefening 51

Het IQ van 15 studenten van de faculteit Communicatiewetenschappen van een universiteit werd gemeten. Dat gaf de volgende resultaten: 98, 90, 91, 95, 105, 102, 100, 93, 94, 103, 99, 97, 100, 104 en 101.

Als we weten dat het gemiddelde IQ van een student van deze universiteit gelijk is aan 105 en σ gelijk is aan 10, kunnen we dan met een foutenrisico α = 5 % besluiten dat studenten van de faculteit Communicatiewetenschappen een lager IQ hebben dan het gemiddelde IQ van studenten aan deze universiteit? Los op met de Z. TEST.

Oefening 52

Een textielbedrijf voor herenkleding maakt jeansbroeken die in maat 50 een lengte hebben van 110 cm. Een klant klaagt dat de broeken te lang zijn en dat op basis van een steekproef van 10 metingen, een gemiddelde van 112 cm en een steekproefstandaardafwijking van 2,5 cm.

Is de klacht gegrond bij een foutenrisico van α = 0,05?

Oefening 53

Een vastgoedmakelaar neemt in een residentiële wijk van Brussel een steekproef van 12 woningen en stelt vast dat de gemiddelde geschatte waarde van een woning gelijk is aan € 780 000 en dat de standaardafwijking gelijk is aan € 49 000.

Test de hypothese dat voor alle woningen in deze wijk de gemiddelde geschatte waarde gelijk is aan € 825 000 tegen de alternatieve hypothese die stelt dat de gemiddelde waarde kleiner is. Gebruik een foutenrisico van 5 %.

Oefening 54

De personeelsdirecteur van een bedrijf beweert dat het personeel te kampen heeft met een gemiddeld overgewicht van 10 kg. Ter natrekking van deze stelling worden 18 personeelsleden willekeurig geselecteerd en onderzocht. Het vastgestelde overgewicht bedraagt 12,4 kg en de steekproefstandaardafwijking is 2,7 kg.

Kan men met een foutenrisico van α = 0,01 de geldigheid van μ gelijk aan 10 kg in twijfel trekken?

Oefening 55

De standaarddeviatie van een normale verdeling wordt verondersteld gelijk te zijn aan 50. Een steekproef van 30 eenheden levert een steekproefstandaardafwijking op van 57.

Kunnen we met een foutenrisico van $\alpha = 0{,}05$ de nulhypothese verwerpen die stelt dat de echte standaarddeviatie gelijk is aan 50?

Oefening 56

Een fabrikant van telescopen staat erop dat zijn toestellen een standaarddeviatie in resolutie hebben van minder dan 2 wanneer ze zijn ingesteld op objecten die 500 lichtjaren verwijderd zijn. Wanneer een nieuwe telescoop 30 keer wordt gebruikt om ingesteld te worden op een object dat 500 lichtjaren verwijderd is, bedraagt de steekproefstandaardafwijking 1,46.

Zou de telescoop verkocht worden? Test H_0 en H_1 bij $\alpha = 0{,}01$.

Oefening 57

Een reclamebureau probeert voor een bepaald product te achterhalen of de interesse afhankelijk is van de leeftijdscategorie waartoe een consument behoort.
Willekeurig werden 375 mensen geselecteerd in vijf vooraf bepaalde leeftijdsgroepen. Het product werd aan iedereen voorgesteld. De resultaten van het onderzoek zijn:

Aankoop van het product in de toekomst	**Leeftijdsgroepen**				
	16 – 28	29 – 42	43 – 55	56 – 68	69 – 82
Frequent	12	18	17	22	32
Zelden	18	25	29	24	30
Nooit	45	32	29	29	13

Gevraagd:

> stel een tabel op met de waargenomen en verwachte frequenties;

> formuleer de nulhypothese en de alternatieve hypothese;

> als $\alpha = 0{,}01$, wordt H_0 dan verworpen of niet?

Hoofdstuk 15

Oefening 58

Een medewerker van de studiedienst van een vakbond werd gevraagd te controleren of het gemiddelde aantal gewerkte uren per week in fabriek A groter is dan het aantal uren in fabriek B.

Het gemiddelde aantal gewerkte uren per week in fabriek A, vastgesteld op basis van een lukrake steekproef van 62 werknemers, bedraagt 43,7 uren. Het gemiddelde aantal uren gewerkt per week in fabriek B is op basis van een lukrake steekproef van 53 werknemers bepaald op 41 uren. De populatiestandaarddeviaties zijn in fabriek A en fabriek B respectievelijk 6 en 5,1 uren.

Test de hypothese dat de arbeiders in fabriek A gemiddeld per week niet langer werken dan de arbeiders in fabriek B. Gebruik een foutenrisico α gelijk aan 5 %.

Oefening 59

Voor een steekproef van 12 jongens en een even grote steekproef van 12 meisjes waren de resultaten (op 100 punten) op een spellingtest Engels de volgende:

Jongens	Meisjes
60	59
82	80
69	48
93	90
85	75
63	60
74	71
79	82
80	75
66	67
74	69
67	61

De populatievarianties zijn respectievelijk 40,96 en 81. Kunnen we met een foutenrisico van α gelijk aan 0,02 besluiten dat de jongens voor de spellingtest beter scoren dan de meisjes? Los op met de Z-toets: twee steekproeven voor gemiddelden.

Oefening 60

Een lukrake steekproef van de wedden van 10 mannelijke bedienden leverde een gemiddelde op van € 40 000 per jaar. Een lukrake steekproef van de wedden van 5 vrouwelijke bedienden leverde een gemiddelde op van € 27 500 per jaar. De gecombineerde schatter s_G^2 was € 100 000 000.

Kunnen we met een foutenrisico van $\alpha = 0{,}05$ besluiten dat er tussen de weddegemiddelden van de mannelijke en vrouwelijke bedienden geen verschil is?

Oefening 61

Een verzekeringsinstelling ontwikkelde een nieuwe trainingsmethode voor het commerciële personeel dat verzekeringscontracten moet verkopen. De onderneming nam een steekproef van 15 bedienden en berekende dat de gemiddelde dagelijkse verkoop van verzekeringscontracten € 2 500 bedroeg en dat s gelijk was aan € 80.
Bij een dochtermaatschappij waar de commerciële bedienden nog steeds de oude commerciële techniek toepassen, werd een steekproef genomen van 16 bedienden. De gemiddelde dagelijkse verkoop was € 2 250 en s bedroeg € 100.
Kan de verzekeringsmaatschappij met een foutenrisico van 2,5 % besluiten dat de gemiddelde dagelijkse verkoop is toegenomen door het nieuwe trainingsprogramma?

Oefening 62

Een steekproef van 8 obligaties en een van 8 aandelen leveren in een bepaald jaar de volgende rendementen op (in %):

Obligaties	7	6	8	9	7	5	8	6
Aandelen	5	11	8	10	7	15	4	6

Kunnen we met een foutenrisico van 5 % besluiten dat het gemiddelde rendement op obligaties lager is dan het gemiddelde rendement op aandelen? Gebruik de T. TEST.

Oefening 63

Los Oefening 62 op met de gegevensanalysefunctie T-toets.

Oefening 64

Van een steekproef van 6 vrouwen die hadden deelgenomen aan een dieetkuur werd het gewicht (in kg) gemeten vóór en na de kuur:

Voor	83	96	97	110	130	96
Na	74	87	74	96	103	82

Als we een foutenrisico van 2 % in acht nemen, kunnen we dan besluiten dat de dieetkuur het gewicht van de vrouwen heeft doen afnemen?

Oefening 65

De heer Roose is productiechef van een assemblagelijn van diskdrives bij nv Tech. De onderneming abonneerde zich op een achtergrondmuziekservice, met de bedoeling de arbeiders meer ontspannen te maken, en daardoor productiever.
De heer Roose staat sceptisch tegenover deze hypothese en denkt dat de muziek zal storen, en zo zal leiden tot een lagere productiviteit. Hij registreerde de wekelijkse productie van dezelfde zes arbeiders vóór de muziekservice inging en nadat de installatie werkte. De onderstaande tabel geeft de geregistreerde gegevens weer:

Arbeider	**1**	**2**	**3**	**4**	**5**	**6**
Werk zonder muziek	219	205	226	198	209	216
Werk met muziek	235	186	240	203	221	205

Kunnen we, met een foutenrisico van $\alpha = 0{,}025$, besluiten dat de productiviteit van de arbeiders verbeterd is? Geef de oplossing met de T. TEST.

Oefening 66

Los Oefening 65 op met de gegevensanalysefunctie T-toets: twee gepaarde steekproeven voor gemiddelden.

Oefening 67

Een leraar economie heeft twee groepen A en B. Groep A telt 16 studenten en groep B 25. Op eenzelfde examen, waarvan de gemiddelde uitslag in A en B weinig verschilde, had groep A een steekproefstandaardafwijking van 9. Bij groep B bedroeg die 12.

Als we een foutenrisico van 5 % in acht nemen, kunnen we dan besluiten dat de variabiliteit van groep B groter is dan die van groep A?

Oefening 68

Een verzekeringsmaatschappij is geïnteresseerd in het aantal hospitalisatiedagen voor diverse ziekten. De maatschappij selecteerde 10 patiënten van ziekenhuis A en 12 van ziekenhuis B, die voor eenzelfde ziekte werden behandeld. De doorgebrachte tijden van de patiënten in ziekenhuis A waren: 2, 3, 4, 1, 4, 2, 8, 2, 7 en 1 dagen. De patiënten van ziekenhuis B bleven 5, 4, 3, 7, 9, 3, 4, 6, 8, 2, 5 en 7 dagen.

Is de variabiliteit in hersteltijd minder variabel voor de patiënten uit ziekenhuis B? Test deze hypothese bij een foutenrisico $\alpha = 0{,}05$. Gebruik de F. TEST.

Oefening 69

Bereken de F-waarde en F_k uit Oefening 64 en stel vast dat je conclusie juist was.

Oefening 70

De nv Biocare is ongerust over de variabiliteit van het aantal bacteriën geproduceerd door twee verschillende culturen. Als de culturen een beduidend verschillende variabiliteit hebben in het aantal voortgebrachte bacteriën, dan zijn de experimenten verknoeid en hebben zich vreemde dingen voorgedaan.

De volgende gegevens werden verzameld:

	Aantal bacteriën (000)										
Cultuur A	91	89	83	101	93	98	144	118	108	125	138
Cultuur B	62	76	90	75	88	99	110	140	145	130	110

Test de nulhypothese dat er geen verschil is in variabiliteit bij een foutenrisico $\alpha = 0{,}02$. Gebruik de F-toets voor twee steekproeven.

Hoofdstuk 16

Oefening 71

In een groot, middelgroot en klein bedrijf wordt uit het bediendepersoneel telkens een steekproef van 5 personeelsleden getrokken. Dat leverde de volgende maandinkomens op (in euro):

Groot bedrijf	Middelgroot bedrijf	Klein bedrijf
3 000	2 250	2 000
2 750	3 000	2 500
3 250	2 000	2 250
4 000	2 500	1 750
5 500	4 000	3 000

Ga na of de verschillen in maandloon al of niet aan het toeval te wijten zijn.

Oefening 72

De onderstaande tabel toont de opbrengsten in bundels per akker van een tarwevariëteit. De akkers werden behandeld met een scheikundige stof A, B of C:

A	B	C
48	47	49
49	49	51
50	48	50
49	48	50

Kunnen we de nulhypothese van dezelfde populatiegemiddelden verwerpen bij een foutenrisico van 5 %?

Oefening 73

Drie arbeidsters in een fabriek verpakken wafels in dozen. Het aantal dozen dat ieder van hen gevuld heeft in drie geselecteerde perioden is opgegeven in de onderstaande tabel:

Periode	Arbeidster A	Arbeidster B	Arbeidster C
8u – 9u	24	18	21
13u – 14u	23	25	16
16u – 17u	25	20	18

Als we een foutenrisico van $\alpha = 5$ % nemen, kunnen we dan besluiten dat er een beduidend verschil is in aantal gevulde dozen per uur, te wijten aan de arbeidsters en dat er ook een beduidend verschil is in aantal gevulde dozen per uur te wijten aan het tijdstip waarop de dozen gevuld werden?

Oefening 74

Een leraar wil drie onderwijsmethoden I, II en III uittesten. Om dat te doen, werden 5 studenten willekeurig gekozen en wordt iedere student onderwezen volgens de verschillende methoden. De leraar wil nu bepalen of er een verschil is tussen de studenten en of er een verschil is tussen de onderwijsmethoden. Nadat ze onderwezen werden volgens de verschillende methoden, nam men van elke student eenzelfde examen af. De resultaten vind je in de onderstaande tabel.

		Methode		
		I	II	III
Student	1	75	81	78
	2	62	85	79
	3	71	68	73
	4	58	92	90
	5	73	90	85

Analyseer het experiment bij een foutenrisico α = 5 %.

Hoofdstuk 17

Oefening 75

Een bierproducent bestudeert het effect van zijn jongste advertentiecampagne. Die hield in dat men willekeurig mensen selecteerde en vroeg hoeveel blikjes bier zij in de voorbije week hadden gekocht en hoeveel advertenties van het bier zij hadden gelezen of gezien in de voorbije week:

X = aantal advertenties
Y = aantal aangekochte blikjes

X	5	8	3	3	1	5	2	3
Y	12	19	10	9	8	7	4	9

> Bereken de regressie aan de hand van de functie LIJNSCH.
> Bereken F_k voor de aanpassing van de regressielijn als α = 0,05. Wat is je conclusie?
> Test de populatieregressiecoëfficiënt A als α = 0,05.
> Bereken met de functie 'Voorspellen' $\hat{y}$ voor de volgende X-waarden: 7, 4 en 6.

Oefening 76

De onderstaande tabel toont achtereenvolgens de lengten x en y van een steekproef van 10 moeders en hun oudste dochters.

Lengte x van de moeder (cm)	166	163	168	161	170	175	165	159	172	162
Lengte y van de oudste dochter (cm)	170	165	170	162	172	169	170	162	177	167

> Teken een spreidingsdiagram.
> Teken de regressielijn.
> Bereken de richtingscoëfficiënt, de constante en s_e.
> Bereken de $\hat{y}$-waarden voor alle x-waarden.

Oefening 77

Een willekeurige steekproef van 8 gezinnen leverde de volgende jaarlijkse inkomens en spaargelden op:

Gezin	Inkomen X (000 euro's)	Spaargelden (000 euro's)
1	22	2,0
2	18	1,8
3	17	1,9
4	27	2,5
5	30	2,0
6	25	2,2
7	20	1,7
8	23	2,1

Pas de gegevensanalysefunctie Regressie toe en bespreek de *Variantieanalysetabel*, de *Coëfficiëntentabel*, de tabel *Storingen*, de tabel *Kans* en de grafische uitvoer.

Oefening 78

Mevrouw De Bondt wil haar woning verkopen. Om de verkoopprijs te bepalen, verzamelde ze informatie over 12 recente woningverkopen. Ze tekende de volgende variabelen op: verkoopprijs (in 000 euro's), aantal m^2 van de woning (in 100 m^2), aantal verdiepingen en de ouderdom van de woning (in jaren):

Y (verkoopprijs)	X_1 (m^2)	X_2 (verdiepingen)	X_3 (ouderdom)
49,65	8,9	1	2
67,95	9,5	1	6
81,15	12,6	2	11
81,60	12,9	2	8
91,50	19,0	2	22
95,25	17,6	1	17
100,35	20,0	2	12
104,25	20,6	2	11
112,65	20,5	1	9
149,70	25,1	2	8
160,65	22,7	2	18
232,50	40,8	3	12

> Bereken de geschatte waarden van y.
> Bereken met LIJNSCH de regressievergelijking.
> Bereken F_k en ga na of de uitkomst een bewijs oplevert tegen of voor H_0.
> Als de woning van mevrouw De Bondt 1 800 m^2 groot is, 1 verdieping heeft en 6 jaar oud is, welke prijs mag zij dan verwachten?

Oefening 79

De nv Staalunie onderzocht de factoren die een invloed hebben op de jaarlijkse verkoop van staal (uitgedrukt in aantal miljoen ton per jaar). Het management denkt dat de volgende factoren van belang zijn: de jaarlijkse inflatievoet, het gemiddelde bedrag per ton waarmee ingevoerd staal goedkoper is dan dat van de Staalunie (in euro) en het aantal voertuigen (in mio) dat wagenproducenten jaarlijks plannen te verkopen (in mio). Er werden gegevens verzameld van 7 opeenvolgende jaren:

Jaar	Y (mio ton verkocht)	X1 (inflatievoet)	X2 (geïmp. prijsreductie)	X3 (aantal wagens)
1	4,2	3,1	3,1	6,2
2	3,1	3,9	5,0	5,1
3	4,0	4,0	2,2	5,7
4	5,0	5,0	2,0	7,1
5	4,3	7,5	2,5	6,5
6	4,1	10,0	3,0	6,1
7	3,7	11,0	3,05	5,9

Gebruik de gegevensanalysefunctie Regressie en bepaal:

> de regressievergelijking;
> het percentage van de totale variatie in het aantal miljoenen ton staal jaarlijks verkocht door de nv Staalunie, verklaard door deze regressievergelijking;
> hoeveel ton staal de vennootschap mag verwachten om te verkopen als de inflatievoet 8,1 % is, de geplande autoverkoop 6 miljoen wagens is en het gemiddelde prijsvoordeel per geïmporteerde ton 3,5 euro is?

Hoofdstuk 18

Oefening 80

De gemiddelde prijzen van aandelen en obligaties genoteerd op een publieke beurs tussen 20X1 en 2X10 zijn gegeven in onderstaande tabel:

Jaar	20X1	20X2	20X3	20X4	20X5	20X6	20X7	20X8	20X9	2X10
Gemiddelde aandelenkoers (euro)	35,22	39,87	41,85	43,23	40,06	53,29	54,14	49,12	40,71	55,15
Gemiddelde obligatiekoers (euro)	102,43	100,93	97,43	97,81	98,32	100,07	97,08	91,59	94,85	94,65

> Teken het spreidingsdiagram en de regressielijn.
> Bepaal de regressievergelijking.
> Bereken de correlatiecoëfficiënt en de determinatiecoëfficiënt.
> Test de nulhypothese dat $\rho \geq 0$ als $\alpha = 0,05$.

Oefening 81

Een producent van voedsel voor koeien wil graag weten welke correlaties er zijn tussen de volgende variabelen:

> de leeftijd van het kalf op het moment dat het een nieuw voedingssupplement wordt toegediend;
> het aanvangsgewicht van het kalf;
> de gewichtstoename, na 1 week voedingssupplement te hebben gekregen.

De volgende gegevens waren het resultaat van een onderzoek bij 8 kalveren:

Kalf	X_1 (leeftijd in weken)	X_2 (begingewicht in kg)	X_3 (gewichtstoename in kg)
1	8	39	7
2	6	52	6
3	7	49	8
4	12	46	10
5	9	61	9
6	6	35	5
7	7	25	3
8	4	55	4

Oefening 82

Een steekproef in 6 landen levert de volgende cijfers op voor x = jaarlijkse sigarettenconsumptie per inwoner en y = jaarlijks dodencijfer door longkankers per 100 000 inwoners:

Land	X	Y
1	3 400	24
2	2 600	20
3	2 200	17
4	2 400	19
5	2 900	26
6	2 100	20

> Bereken r.
> Bij een foutenrisico α = 0,025 kunnen we de nulhypothese verwerpen.
> Test of er bij een foutenrisico α = 0,05 een relatie bestaat tussen de sigarettenconsumptie en longkanker.

Hoofdstuk 19

Oefening 83

De onderstaande tabel geeft de gemiddelde maandelijkse productie van een grondstof in miljoen kg voor de jaren 20X5 – 2X15.

Bereken een vier jaar voortschrijdend gemiddelde en teken een grafiek van de trendwaarden en de werkelijke waarden.

Tabel

Jaartal	20X5	20X6	20X7	20X8	20X9	2X10	2X11	2X12	2X13	2X14	2X15
Gemiddelde maandelijkse productie van de grondstof (in mio kg)	5	3,65	4,30	4,45	3,89	3,81	3,26	3,87	4,17	4,11	3,38

Oefening 84

Om de verkoop van brood de volgende maand te kunnen voorspellen, gebruikte een bakkerij een exponentiële afvlakkingsformule met $\alpha = 0{,}40$:

$$T_{t+1} = 0{,}40T_t + (1 - 0{,}40) \cdot Y_t$$

De eerste maand van het jaar werden 24 700 broden effectief verkocht, de volgende maanden waren de verkopen:

26 000 – 24 800 – 23 700 – 23 600 – 23 100 – 23 000 – 23 080 – 23 250 – 23 900 – 24 350 – 25 850

> Bereken de maandelijkse trendwaarden en zet de trend- en werkelijke verkopen in een grafiek.
> Bereken ook de verkoopsprognose voor de eerste maand van het volgende jaar.

Oefening 85

Onderstaande tabel geeft de productie van sigaretten in een land gedurende de jaren 20X0 - 2X15 weer:

Jaar	Hoeveelheid (mio)	Jaar	Hoeveelheid (mio)
20X0	47,1	20X8	47,2
20X1	46,1	20X9	47,1
20X2	47,0	2X10	48,0
20X3	47,3	2X11	48,0
20X4	47,1	2X12	48,1
20X5	47,0	2X13	46,9
20X6	47,5	2X14	47,9
20X7	47,6	2X15	49,0

> Bepaal de trendvergelijking en teken de grafiek.
> Schat de jaarlijkse productie voor de jaren 2X16 en 2X17.

Oefening 86

Volgende tabel geeft het verloop van het aantal arbeiders in de industrie in een bepaald land weer gedurende de jaren 20X1-20X9.

Jaar	20X1	20X2	20X3	20X4	20X5	20X6	20X7	20X8	20X9
Arbeiders (mio)	9,96	9,93	9,55	9,15	8,86	8,64	8,36	7,82	7,58

> Bepaal de trendvergelijking en teken de grafiek.
> Bereken de trendwaarden.
> Schat het aantal arbeiders in het jaar 20X0.
> Schat het aantal arbeiders in het jaar 2X10.

Bijlage 2 Lijst van de gebruikte werkbladfuncties

Naam	Functie	Engelse functienaam
AANTAL	telt het aantal cellen dat getallen bevat	COUNT
AANTAL. ALS	telt het aantal cellen dat in een bereik voldoet aan een opgegeven criterium	COUNTIF
AANTAL. LEGE. CELLEN	telt het aantal lege cellen in een bereik	COUNTBLANK
AANTALARG	telt het aantal cellen dat tekst en getallen bevat in een gegevenslijst	COUNTA
BINOMIALE. INV	berekent de kritieke waarde van X in een hypothesetest gebaseerd op een binomiale verdeling	BINOM. INV
BINOMIALE. VERD	geeft als resultaat de binomiale verdeling	BINOMDIST
CHIKW. TEST	berekent het resultaat van een onafhankelijkheidstoets	CHISQ. TEST
CHIKW. INV	berekent de inverse van een eenzijdige kans van de chikwadraatverdeling	CHISQ. INV
CHIKW. VERD	berekent de linkszijdige kans van de chikwadraatverdeling	CHISQ. DIST
COMBINATIES	berekent het aantal combinaties van n objecten in groepen van r objecten	COMBIN
CORRELATIE	berekent de correlatiecoëfficient van twee gegevensverzamelingen	CORREL
COVARIANTIE. S	berekent de covariantie voor een steekproef	COVARIANCE. S
FACULTEIT	berekent de faculteit van een getal	FACT
F. INV	berekent de inverse van de F-verdeling	F. INV
F. INV. RECHTS	met deze functie vergelijkt men de mate van variabiliteit in twee gegevensverzamelingen	F. INV. RT
F. TEST	berekent de tweezijdige kans van F om te bepalen of de varianties van twee steekproeven van elkaar verschillen	F. TEST
F. VERD	gebruik je om uit te maken of de gegevens in twee verzamelingen in verschillende mate variëren	F. DIST
GEMIDDELDE	berekent het rekenkundig gemiddelde van getallen in een verzameling	AVERAGE
GETRIMD. GEM	berekent het rekenkundig gemiddelde van getallen in een verzameling nadat extreme getallen werden uitgesloten	TRIMMEAN
GROOTSTE	berekent de op k – 1 na grootste waarde in een gegevensbereik	LARGE

HARM. GEM.	berekent het harmonisch gemiddelde van een verzameling getallen	HARM. MEAN
INTERVAL	berekent hoe vaak waarden voorkomen in een waardenbereik	FREQUENCY
KLEINSTE	berekent de op k – 1 na kleinste waarde in een gegevensbereik	SMALL
KWARTIEL. INC	berekent een kwartiel van een gegevens-verzameling	QUARTILE. INC
LIJNSCH	geeft als resultaat een matrix die een rechte lijn beschrijft die het best past bij de gegevens	LINEST
MAX	geeft als resultaat de grootste waarde in een lijst met waarden	MAX
MEDIAAN	berekent de middelste waarde van de gegeven getallen	MEDIAN
MEETK. GEM	geeft als resultaat het meetkundig gemiddelde van positieve numerieke gegevens in een matrix	GEOMEAN
MIN	geeft als resultaat het kleinste getal in een lijst met waarden	MIN
MODUS. ENKELV	geeft als resultaat de meest voorkomende waarde in een bereik met gegevens	MODE. SNGL
MODUS. MEERV	geeft alls resultaat een verticale matrix met de waarden uit een gegevensverzameling die de hoogste frequentie hebben	MODE. MULT
NORM. INV. N	berekent voor een opgegeven μ en σ de waarde X waarvan de kans dat die voorkomt of kleiner is, gelijk is aan een opgegeven kans	NORM. INV
NORM. S. INV	berekent voor μ gelijk aan nul en een σ gelijk aan 1 de waarde Z waarvan de kans dat die voorkomt of kleiner is, gelijk is aan een opgegeven kans	NORM. S. INV
NORM. S. VERD	geeft als resultaat de normale standaard-verdeling	NORM. S. DIST
NORM. VERD. N	berekent de cumulatieve normale verdeling of de kans dat X kleiner of gelijk is aan een bepaalde waarde voor een opgegeven μ en σ	NORM. DIST
NORMALISEREN	berekent een genormaliseerde waarde uit een verdeling	STANDARDIZE
PEARSON	berekent de correlatiecoëfficiënt r van Pearson	PEARSON
PROCENTRANG. INC	berekent de relatieve positie van een waarde in een gegevensbereik	PERCENTRANK. INC
PERCENTIEL. INC	berekent het k-de percentiel van waarden in een bereik	PERCENTILE. INC
PERMUTATIES	berekent het aantal permutaties voor een gegeven aantal objecten dat uit het totale aantal objecten geselecteerd kan worden	PERMUT

POISSON. VERD	geeft als resultaat de Poisson-verdeling	POISSON. DIST
RANG. GELIJK	berekent de rang van een getal in een lijst getallen	RANK. EQ
RICHTING	berekent de richtingscoëfficiënt van een lineaire regressielijn	SLOPE
R. KWADRAAT	berekent de determinatiecoëfficiënt van een lineaire regressielijn	RSQ
SCHEEFHEID	berekent de mate van asymmetrie van een verdeling	SKEW
SNIJPUNT	berekent het snijpunt van een lijn met de Y-as aan de hand van een optimale regressielijn	INTERCEPT
SOM	telt de getallen in een cellenbereik op	SUM
STAND. FOUT. YX	berekent de standaardafwijking van de Y-waarden rond de geschatte Y-waarden	STEYX
STDEV. S	maakt een schatting van de standaarddeviatie op basis van een steekproef	STDEV. S
STDEV. P	berekent de standaarddeviatie van een volledige populatie	STDEV. P
T. INV	berekent de t-waarde van een t-verdeling als functie van een opgegeven kans en een aantal opgegeven vrijheidsgraden	T. INV
T. INV. 2T	berekent de t-waarde van de t-verdeling als een functie van een opgegeven kans en een opgegeven aantal vrijheidsgraden die verband houden met een tweezijdige t-verdeling	T. INV. 2T
T. TEST	berekent de kans dat twee steekproeven behoren tot dezelfde twee onderliggende populaties met hetzelfde gemiddelde	T. TEST
T. DIST	berekent de eenzijdige cumulatieve t-verdelings-functie	T. DIST
TREND	berekent geschatte waarden van y voor waar-genomen x-waarden en nieuwe x-waarden	TREND
VAR. P	berekent de variantie op basis van een volledige populatie	VAR. P
VAR. S	maakt een schatting van de variantie op basis van een steekproef	VAR. S
VERTROUWELIJKHEID. NORM	berekent de steekproeffout voor een gemiddelde waarde van de elementen van een populatie met een normale verdeling	CONFIDENCE. NORM
VERTROUWELIJKHEID. T	berekent de steekproeffout voor een gemiddelde waarde van de elementen van een populatie met behulp van een student t-verdeling	CONFIDENCE. T
VOORSPELLEN. LINEAIR	berekent aan de hand van bestaande waarden een toekomstige waarde volgens een lineaire trend	FORECAST. LINEAR
Z. TEST	berekent de eenzijdige P-waarde van een Z-toets	Z.TEST

Bijlage 3 Lijst van de gebruikte gegevensanalysefuncties

Naam	Functie	Engelse functienaam
Beschrijvende statistiek	deze functie levert een univariaat statistisch rapport voor gegevens in een invoerbereik. Het rapport geeft informatie over de centrale trend en de variabiliteit van de gegevens	Descriptive statistics
Correlatie	berekent meerdere samenhangen indien het aantal gegevensverzamelingen groter is dan twee	Correlation
Exponentiële demping	hiermee wordt een waarde voorspeld op basis van de trendwaarde en de werkelijke waarde van de vorige periode	Exponential smoothing
F-toets voor twee steekproeven	gebruikt om populatievarianties van twee steekproeven te testen	F-test Two Sample for Variances
Histogram	hiermee kun je afzonderlijke en cumulatieve frequenties berekenen voor een cellenbereik met gegevens en grenswaarden	Histogram
Multifactoriële variantieanalyse zonder herhaling	met deze functie voer je een multifactoriële variantieanalyse uit waarbij je de hypothese test dat gemiddelden van twee of meer steekproeven gerangschikt volgens meerdere factoren gelijk zijn	Anova: Two-Factor Without Replication
Rang en percentiel	berekent van ieder gegeven uit een gegevensverzameling tegelijkertijd de rang en het percentiel	Rank and Percentile
Regressie	hiermee pas je een lineaire regressieanalyse toe voor een afhankelijke variabele en een of meer onafhankelijke variabelen	Regression
T-test: twee gepaarde steekproeven voor gemiddelden	gebruikt wanneer er sprake is van een natuurlijke gepaardheid van waarnemingen in de steekproeven	t-Test: Paired Two Sample for Means
T-toets: twee steekproeven met ongelijke varianties	gebruik je om te bepalen of de gemiddelden van twee steekproeven gelijk zijn	t-Test: Two Sample Assuming Unequal Variances

Unifactoriële variantieanalyse	met deze functie kun je een eenvoudige variantieanalyse uitvoeren, waarmee de hypothese dat gemiddelden van verschillende steekproeven gelijk zijn, kan worden getest	Anova: Single Factor
Z-toets: twee steek-proeven voor gemiddelden	gebruikt om hypothesen over het verschil van twee populatiegemiddelden te testen	Z-Test: Two Sample for Means
Zwevend gemiddelde	berekent een reeks van voortschrijdende gemiddelden	Moving Average

Bijlage 4 Tabellen

Tabel 1 De individuele binomiale kansen P(X)

		p										
n	**x**	**0.1**	**0.2**	**0.25**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.75**	**0.8**	**0.9**
1	0	0.900	0.800	0.750	0.700	0.600	0.500	0.400	0.300	0.250	0.200	0.100
	1	0.100	0.200	0.250	0.300	0.400	0.500	0.600	0.700	0.750	0.800	0.900
2	0	0.810	0.640	0.563	0.490	0.360	0.250	0.160	0.090	0.063	0.040	0.010
	1	0.180	0.320	0.375	0.420	0.480	0.500	0.480	0.420	0.375	0.320	0.180
	2	0.010	0.040	0.063	0.090	0.160	0.250	0.360	0.490	0.563	0.640	0.810
3	0	0.729	0.512	0.422	0.343	0.216	0.125	0.064	0.027	0.016	0.008	0.001
	1	0.243	0.384	0.422	0.441	0.432	0.375	0.288	0.189	0.141	0.096	0.027
	2	0.027	0.096	0.141	0.189	0.288	0.375	0.432	0.441	0.422	0.384	0.243
	3	0.001	0.008	0.016	0.027	0.064	0.125	0.216	0.343	0.422	0.512	0.729
4	0	0.656	0.410	0.316	0.240	0.130	0.063	0.026	0.008	0.004	0.002	0.000
	1	0.292	0.410	0.422	0.412	0.346	0.250	0.154	0.076	0.047	0.026	0.004
	2	0.049	0.154	0.211	0.265	0.346	0.375	0.346	0.265	0.211	0.154	0.049
	3	0.004	0.026	0.047	0.076	0.154	0.250	0.346	0.412	0.422	0.410	0.292
	4	0.000	0.002	0.004	0.008	0.026	0.063	0.130	0.240	0.316	0.410	0.656
5	0	0.590	0.328	0.237	0.168	0.078	0.031	0.010	0.002	0.001	0.000	0.000
	1	0.328	0.410	0.396	0.360	0.259	0.156	0.077	0.028	0.015	0.006	0.000
	2	0.073	0.205	0.264	0.309	0.346	0.312	0.230	0.132	0.088	0.051	0.008
	3	0.008	0.051	0.088	0.132	0.230	0.312	0.346	0.309	0.264	0.205	0.073
	4	0.000	0.006	0.015	0.028	0.077	0.156	0.259	0.360	0.396	0.410	0.328
	5	0.000	0.000	0.001	0.002	0.010	0.031	0.078	0.168	0.237	0.328	0.590
6	0	0.531	0.262	0.178	0.118	0.047	0.016	0.004	0.001	0.000	0.000	0.000
	1	0.354	0.393	0.356	0.303	0.187	0.094	0.037	0.010	0.004	0.002	0.000
	2	0.098	0.246	0.297	0.324	0.311	0.234	0.138	0.060	0.033	0.015	0.001
	3	0.015	0.082	0.132	0.185	0.276	0.313	0.276	0.185	0.132	0.082	0.015
	4	0.001	0.015	0.033	0.060	0.138	0.234	0.311	0.324	0.297	0.246	0.098
	5	0.000	0.002	0.004	0.010	0.037	0.094	0.187	0.303	0.356	0.393	0.354
	6	0.000	0.000	0.000	0.001	0.004	0.016	0.047	0.118	0.178	0.262	0.531
7	0	0.478	0.210	0.133	0.082	0.028	0.008	0.002	0.000	0.000	0.000	0.000
	1	0.372	0.367	0.311	0.247	0.131	0.055	0.017	0.004	0.001	0.000	0.000
	2	0.124	0.275	0.311	0.318	0.261	0.164	0.077	0.025	0.012	0.004	0.000
	3	0.023	0.115	0.173	0.227	0.290	0.273	0.194	0.097	0.058	0.029	0.003
	4	0.003	0.029	0.058	0.097	0.194	0.273	0.290	0.227	0.173	0.115	0.023
	5	0.000	0.004	0.012	0.025	0.077	0.164	0.261	0.318	0.311	0.275	0.124
	6	0.000	0.000	0.001	0.004	0.017	0.055	0.131	0.247	0.311	0.367	0.372
	7	0.000	0.000	0.000	0.000	0.002	0.008	0.028	0.082	0.133	0.210	0.478

		p										
n	x	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9
8	0	0.430	0.168	0.100	0.058	0.017	0.004	0.001	0.000	0.000	0.000	0.000
	1	0.383	0.336	0.267	0.198	0.090	0.031	0.008	0.001	0.000	0.000	0.000
	2	0.149	0.294	0.311	0.296	0.209	0.109	0.041	0.010	0.004	0.001	0.000
	3	0.033	0.147	0.208	0.254	0.279	0.219	0.124	0.047	0.023	0.009	0.000
	4	0.005	0.046	0.087	0.136	0.232	0.273	0.232	0.136	0.087	0.046	0.005
	5	0.000	0.009	0.023	0.047	0.124	0.219	0.279	0.254	0.208	0.147	0.033
	6	0.000	0.001	0.004	0.010	0.041	0.109	0.209	0.296	0.311	0.294	0.149
	7	0.000	0.000	0.000	0.001	0.008	0.031	0.090	0.198	0.267	0.336	0.383
	8	0.000	0.000	0.000	0.000	0.001	0.004	0.017	0.058	0.100	0.168	0.430
9	0	0.387	0.134	0.075	0.040	0.010	0.002	0.000	0.000	0.000	0.000	0.000
	1	0.387	0.302	0.225	0.156	0.060	0.018	0.004	0.000	0.000	0.000	0.000
	2	0.172	0.302	0.300	0.267	0.161	0.070	0.021	0.004	0.001	0.000	0.000
	3	0.045	0.176	0.234	0.267	0.251	0.164	0.074	0.021	0.009	0.003	0.000
	4	0.007	0.066	0.117	0.172	0.251	0.246	0.167	0.074	0.039	0.017	0.001
	5	0.001	0.017	0.039	0.074	0.167	0.246	0.251	0.172	0.117	0.066	0.007
	6	0.000	0.003	0.009	0.021	0.074	0.164	0.251	0.267	0.234	0.176	0.045
	7	0.000	0.000	0.001	0.004	0.021	0.070	0.161	0.267	0.300	0.302	0.172
	8	0.000	0.000	0.000	0.000	0.004	0.018	0.060	0.156	0.225	0.302	0.387
	9	0.000	0.000	0.000	0.000	0.000	0.002	0.010	0.040	0.075	0.134	0.387
10	0	0.349	0.107	0.056	0.028	0.006	0.001	0.000	0.000	0.000	0.000	0.000
	1	0.387	0.268	0.188	0.121	0.040	0.010	0.002	0.000	0.000	0.000	0.000
	2	0.194	0.302	0.282	0.233	0.121	0.044	0.011	0.001	0.000	0.000	0.000
	3	0.057	0.201	0.250	0.267	0.215	0.117	0.042	0.009	0.003	0.001	0.000
	4	0.011	0.088	0.146	0.200	0.251	0.205	0.111	0.037	0.016	0.006	0.000
	5	0.001	0.026	0.058	0.103	0.201	0.246	0.201	0.103	0.058	0.026	0.001
	6	0.000	0.006	0.016	0.037	0.111	0.205	0.251	0.200	0.146	0.088	0.011
	7	0.000	0.001	0.003	0.009	0.042	0.117	0.215	0.267	0.250	0.201	0.057
	8	0.000	0.000	0.000	0.001	0.011	0.044	0.121	0.233	0.282	0.302	0.194
	9	0.000	0.000	0.000	0.000	0.002	0.010	0.040	0.121	0.188	0.268	0.387
	10	0.000	0.000	0.000	0.000	0.000	0.001	0.006	0.028	0.056	0.107	0.349
11	0	0.314	0.086	0.042	0.020	0.004	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.384	0.236	0.155	0.093	0.027	0.005	0.001	0.000	0.000	0.000	0.000
	2	0.213	0.295	0.258	0.200	0.089	0.027	0.005	0.001	0.000	0.000	0.000
	3	0.071	0.221	0.258	0.257	0.177	0.081	0.023	0.004	0.001	0.000	0.000
	4	0.016	0.111	0.172	0.220	0.236	0.161	0.070	0.017	0.006	0.002	0.000
	5	0.002	0.039	0.080	0.132	0.221	0.226	0.147	0.057	0.027	0.010	0.000
	6	0.000	0.010	0.027	0.057	0.147	0.226	0.221	0.132	0.080	0.039	0.002
	7	0.000	0.002	0.006	0.017	0.070	0.161	0.236	0.220	0.172	0.111	0.016
	8	0.000	0.000	0.001	0.004	0.023	0.081	0.177	0.257	0.258	0.221	0.071
	9	0.000	0.000	0.000	0.001	0.005	0.027	0.089	0.200	0.258	0.295	0.213
	10	0.000	0.000	0.000	0.000	0.001	0.005	0.027	0.093	0.155	0.236	0.384
	11	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.020	0.042	0.086	0.314

		p										
n	x	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9
12	0	0.282	0.069	0.032	0.014	0.002	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.377	0.206	0.127	0.071	0.017	0.003	0.000	0.000	0.000	0.000	0.000
	2	0.230	0.283	0.232	0.168	0.064	0.016	0.002	0.000	0.000	0.000	0.000
	3	0.085	0.236	0.258	0.240	0.142	0.054	0.012	0.001	0.000	0.000	0.000
	4	0.021	0.133	0.194	0.231	0.213	0.121	0.042	0.008	0.002	0.001	0.000
	5	0.004	0.053	0.103	0.158	0.227	0.193	0.101	0.029	0.011	0.003	0.000
	6	0.000	0.016	0.040	0.079	0.177	0.226	0.177	0.079	0.040	0.016	0.000
	7	0.000	0.003	0.011	0.029	0.101	0.193	0.227	0.158	0.103	0.053	0.004
	8	0.000	0.001	0.002	0.008	0.042	0.121	0.213	0.231	0.194	0.133	0.021
	9	0.000	0.000	0.000	0.001	0.012	0.054	0.142	0.240	0.258	0.236	0.085
	10	0.000	0.000	0.000	0.000	0.002	0.016	0.064	0.168	0.232	0.283	0.230
	11	0.000	0.000	0.000	0.000	0.000	0.003	0.017	0.071	0.127	0.206	0.377
	12	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.014	0.032	0.069	0.282
13	0	0.254	0.055	0.024	0.010	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.367	0.179	0.103	0.054	0.011	0.002	0.000	0.000	0.000	0.000	0.000
	2	0.245	0.268	0.206	0.139	0.045	0.010	0.001	0.000	0.000	0.000	0.000
	3	0.100	0.246	0.252	0.218	0.111	0.035	0.006	0.001	0.000	0.000	0.000
	4	0.028	0.154	0.210	0.234	0.184	0.087	0.024	0.003	0.001	0.000	0.000
	5	0.006	0.069	0.126	0.180	0.221	0.157	0.066	0.014	0.005	0.001	0.000
	6	0.001	0.023	0.056	0.103	0.197	0.209	0.131	0.044	0.019	0.006	0.000
	7	0.000	0.006	0.019	0.044	0.131	0.209	0.197	0.103	0.056	0.023	0.001
	8	0.000	0.001	0.005	0.014	0.066	0.157	0.221	0.180	0.126	0.069	0.006
	9	0.000	0.000	0.001	0.003	0.024	0.087	0.184	0.234	0.210	0.154	0.028
	10	0.000	0.000	0.000	0.001	0.006	0.035	0.111	0.218	0.252	0.246	0.100
	11	0.000	0.000	0.000	0.000	0.001	0.010	0.045	0.139	0.206	0.268	0.245
	12	0.000	0.000	0.000	0.000	0.000	0.002	0.011	0.054	0.103	0.179	0.367
	13	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.010	0.024	0.055	0.254
14	0	0.229	0.044	0.018	0.007	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.356	0.154	0.083	0.041	0.007	0.001	0.000	0.000	0.000	0.000	0.000
	2	0.257	0.250	0.180	0.113	0.032	0.006	0.001	0.000	0.000	0.000	0.000
	3	0.114	0.250	0.240	0.194	0.085	0.022	0.003	0.000	0.000	0.000	0.000
	4	0.035	0.172	0.220	0.229	0.155	0.061	0.014	0.001	0.000	0.000	0.000
	5	0.008	0.086	0.147	0.196	0.207	0.122	0.041	0.007	0.002	0.000	0.000
	6	0.001	0.032	0.073	0.126	0.207	0.183	0.092	0.023	0.008	0.002	0.000
	7	0.000	0.009	0.028	0.062	0.157	0.209	0.157	0.062	0.028	0.009	0.000
	8	0.000	0.002	0.008	0.023	0.092	0.183	0.207	0.126	0.073	0.032	0.001
	9	0.000	0.000	0.002	0.007	0.041	0.122	0.207	0.196	0.147	0.086	0.008
	10	0.000	0.000	0.000	0.001	0.014	0.061	0.155	0.229	0.220	0.172	0.035
	11	0.000	0.000	0.000	0.000	0.003	0.022	0.085	0.194	0.240	0.250	0.114
	12	0.000	0.000	0.000	0.000	0.001	0.006	0.032	0.113	0.180	0.250	0.257
	13	0.000	0.000	0.000	0.000	0.000	0.001	0.007	0.041	0.083	0.154	0.356
	14	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.007	0.018	0.044	0.229

		p										
n	**x**	**0.1**	**0.2**	**0.25**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.75**	**0.8**	**0.9**
15	0	0.206	0.035	0.013	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.343	0.132	0.067	0.031	0.005	0.000	0.000	0.000	0.000	0.000	0.000
	2	0.267	0.231	0.156	0.092	0.022	0.003	0.000	0.000	0.000	0.000	0.000
	3	0.129	0.250	0.225	0.170	0.063	0.014	0.002	0.000	0.000	0.000	0.000
	4	0.043	0.188	0.225	0.219	0.127	0.042	0.007	0.001	0.000	0.000	0.000
	5	0.010	0.103	0.165	0.206	0.186	0.092	0.024	0.003	0.001	0.000	0.000
	6	0.002	0.043	0.092	0.147	0.207	0.153	0.061	0.012	0.003	0.001	0.000
	7	0.000	0.014	0.039	0.081	0.177	0.196	0.118	0.035	0.013	0.003	0.000
	8	0.000	0.003	0.013	0.035	0.118	0.196	0.177	0.081	0.039	0.014	0.000
	9	0.000	0.001	0.003	0.012	0.061	0.153	0.207	0.147	0.092	0.043	0.002
	10	0.000	0.000	0.001	0.003	0.024	0.092	0.186	0.206	0.165	0.103	0.010
	11	0.000	0.000	0.000	0.001	0.007	0.042	0.127	0.219	0.225	0.188	0.043
	12	0.000	0.000	0.000	0.000	0.002	0.014	0.063	0.170	0.225	0.250	0.129
	13	0.000	0.000	0.000	0.000	0.000	0.003	0.022	0.092	0.156	0.231	0.267
	14	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.031	0.067	0.132	0.343
	15	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.013	0.035	0.206
20	0	0.122	0.012	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.270	0.058	0.021	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	2	0.285	0.137	0.067	0.028	0.003	0.000	0.000	0.000	0.000	0.000	0.000
	3	0.190	0.205	0.134	0.072	0.012	0.001	0.000	0.000	0.000	0.000	0.000
	4	0.090	0.218	0.190	0.130	0.035	0.005	0.000	0.000	0.000	0.000	0.000
	5	0.032	0.175	0.202	0.179	0.075	0.015	0.001	0.000	0.000	0.000	0.000
	6	0.009	0.109	0.169	0.192	0.124	0.037	0.005	0.000	0.000	0.000	0.000
	7	0.002	0.055	0.112	0.164	0.166	0.074	0.015	0.001	0.000	0.000	0.000
	8	0.000	0.022	0.061	0.114	0.180	0.120	0.035	0.004	0.001	0.000	0.000
	9	0.000	0.007	0.027	0.065	0.160	0.160	0.071	0.012	0.003	0.000	0.000
	10	0.000	0.002	0.010	0.031	0.117	0.176	0.117	0.031	0.010	0.002	0.000
	11	0.000	0.000	0.003	0.012	0.071	0.160	0.160	0.065	0.027	0.007	0.000
	12	0.000	0.000	0.001	0.004	0.035	0.120	0.180	0.114	0.061	0.022	0.000
	13	0.000	0.000	0.000	0.001	0.015	0.074	0.166	0.164	0.112	0.055	0.002
	14	0.000	0.000	0.000	0.000	0.005	0.037	0.124	0.192	0.169	0.109	0.009
	15	0.000	0.000	0.000	0.000	0.001	0.015	0.075	0.179	0.202	0.175	0.032
	16	0.000	0.000	0.000	0.000	0.000	0.005	0.035	0.130	0.190	0.218	0.090
	17	0.000	0.000	0.000	0.000	0.000	0.001	0.012	0.072	0.134	0.205	0.190
	18	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.028	0.067	0.137	0.285
	19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.007	0.021	0.058	0.270
	20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.012	0.122

Tabel 2 Binomiaalcoëfficiënten $\binom{n}{x}$

n	X 0	1	2	3	4	5	6	7	8	9	10
0	1										
1	1	1									
2	1	2	1								
3	1	3	3	1							
4	1	4	6	4	1						
5	1	5	10	10	5	1					
6	1	6	15	20	15	6	1				
7	1	7	21	35	35	21	7	1			
8	1	8	28	56	70	56	28	8	1		
9	1	9	36	84	126	125	84	36	9	1	
10	1	10	45	120	210	252	210	120	45	10	1
11	1	11	55	165	330	462	462	330	165	55	11
12	1	12	66	220	495	792	924	792	495	220	66
13	1	13	78	286	715	1.287	1.716	1.716	1.287	715	286
14	1	14	91	364	1.001	2.002	3.003	3.432	3.003	2.002	1.001
15	1	15	105	455	1.365	3.003	5.005	6.435	6.435	5.005	3.003
16	1	16	120	560	1.820	4.368	8.008	11.440	12.870	11.440	8.008
17	1	17	136	680	2.380	6.188	12.376	19.448	24.310	24.310	19.448
18	1	18	153	816	3.060	8.568	18.564	31.824	43.758	48.620	43.758
19	1	19	171	969	3.876	11.628	27.132	50.388	75.582	92.378	92.378
20	1	20	190	1.140	4.845	15.504	38.760	77.520	125.970	167.960	184.756

Tabel 3 Poisson-kansen

(de kans van een bepaalde waarde van X voor een gegeven waarde van λ)

X	λ									
	0,1	**0,2**	**0,3**	**0,4**	**0,5**	**0,6**	**0,7**	**0,8**	**0,9**	**1,0**
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066	0,3679
1	0,0905	0,1637	0,2222	0,2681	0,3033	0,3293	0,3476	0,3595	0,3659	0,3679
2	0,0045	0,0164	0,0333	0,0536	0,0758	0,0988	0,1217	0,1438	0,1647	0,1839
3	0,0002	0,0011	0,0033	0,0072	0,0126	0,0198	0,0284	0,0383	0,0494	0,0613
4	0,0000	0,0001	0,0003	0,0007	0,0016	0,0030	0,0050	0,0077	0,0111	0,0153
5	0,0000	0,0000	0,0000	0,0001	0,0002	0,0004	0,0007	0,0012	0,0020	0,0031
6	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0002	0,0003	0,0005
7	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001
	1,1	**1,2**	**1,3**	**1,4**	**1,5**	**1,6**	**1,7**	**1,8**	**1,9**	**2,0**
0	0,3329	0,3012	0,2725	0,2466	0,2231	0,2019	0,1827	0,1653	0,1496	0,1353
1	0,3662	0,3614	0,3543	0,3452	0,3347	0,3230	0,3106	0,2975	0,2842	0,2707
2	0,2014	0,2169	0,2303	0,2417	0,2510	0,2584	0,2640	0,2678	0,2700	0,2707
3	0,0738	0,0867	0,0998	0,1128	0,1255	0,1378	0,1496	0,1607	0,1710	0,1804
4	0,0203	0,0260	0,0324	0,0395	0,0471	0,0551	0,0636	0,0723	0,0812	0,0902
5	0,0045	0,0062	0,0084	0,0111	0,0141	0,0176	0,0216	0,0260	0,0309	0,0361
6	0,0008	0,0012	0,0018	0,0026	0,0035	0,0047	0,0061	0,0078	0,0098	0,0120
7	0,0001	0,0002	0,0003	0,0005	0,0008	0,0011	0,0015	0,0020	0,0027	0,0034
8	0,0000	0,0000	0,0001	0,0001	0,0001	0,0002	0,0003	0,0005	0,0006	0,0009
9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	0,0001	0,0002
	2,1	**2,2**	**2,3**	**2,4**	**2,5**	**2,6**	**2,7**	**2,8**	**2,9**	**3,0**
0	0,1225	0,1108	0,1003	0,0907	0,0821	0,0743	0,0672	0,0608	0,0550	0,0498
1	0,2572	0,2438	0,2306	0,2177	0,2052	0,1931	0,1815	0,1703	0,1596	0,1494
2	0,2700	0,2681	0,2652	0,2613	0,2565	0,2510	0,2450	0,2384	0,2314	0,2240
3	0,1890	0,1966	0,2033	0,2090	0,2138	0,2176	0,2205	0,2225	0,2237	0,2240
4	0,0992	0,1082	0,1169	0,1254	0,1336	0,1414	0,1488	0,1557	0,1622	0,1680
5	0,0417	0,0476	0,0538	0,0602	0,0668	0,0735	0,0804	0,0872	0,0940	0,1008
6	0,0146	0,0174	0,0206	0,0241	0,0278	0,0319	0,0362	0,0407	0,0455	0,0504
7	0,0044	0,0055	0,0068	0,0083	0,0099	0,0118	0,0139	0,0163	0,0188	0,0216
8	0,0011	0,0015	0,0019	0,0025	0,0031	0,0038	0,0047	0,0057	0,0068	0,0081
9	0,0003	0,0004	0,0005	0,0007	0,0009	0,0011	0,0014	0,0018	0,0022	0,0027
10	0,0001	0,0001	0,0001	0,0002	0,0002	0,0003	0,0004	0,0005	0,0006	0,0008
11	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	0,0001	0,0002	0,0002
12	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001
	3,1	**3,2**	**3,3**	**3,4**	**3,5**	**3,6**	**3,7**	**3,8**	**3,9**	**4,0**
0	0,0450	0,0408	0,0369	0,0334	0,0302	0,0273	0,0247	0,0224	0,0202	0,0183
1	0,1397	0,1304	0,1217	0,1135	0,1057	0,0984	0,0915	0,0850	0,0789	0,0733
2	0,2165	0,2087	0,2008	0,1929	0,1850	0,1771	0,1692	0,1615	0,1539	0,1465
3	0,2237	0,2226	0,2209	0,2186	0,2158	0,2125	0,2087	0,2046	0,2001	0,1954
4	0,1734	0,1781	0,1823	0,1858	0,1888	0,1912	0,1931	0,1944	0,1951	0,1954
5	0,1075	0,1140	0,1203	0,1264	0,1322	0,1377	0,1429	0,1477	0,1522	0,1563
6	0,0555	0,0608	0,0662	0,0716	0,0771	0,0826	0,0881	0,0936	0,0989	0,1042
7	0,0246	0,0278	0,0312	0,0348	0,0385	0,0425	0,0466	0,0508	0,0551	0,0595
8	0,0095	0,0111	0,0129	0,0148	0,0169	0,0191	0,0215	0,0241	0,0269	0,0298
9	0,0033	0,0040	0,0047	0,0056	0,0066	0,0076	0,0089	0,0102	0,0116	0,0132
10	0,0010	0,0013	0,0016	0,0019	0,0023	0,0028	0,0033	0,0039	0,0045	0,0053
11	0,0003	0,0004	0,0005	0,0006	0,0007	0,0009	0,0011	0,0013	0,0016	0,0019
12	0,0001	0,0001	0,0001	0,0002	0,0002	0,0003	0,0003	0,0004	0,0005	0,0006
13	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	0,0001	0,0001	0,0002	0,0002
14	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001

X	λ									
	4,1	**4,2**	**4,3**	**4,4**	**4,5**	**4,6**	**4,7**	**4,8**	**4,9**	**5,0**
0	0,0166	0,0150	0,0136	0,0123	0,0111	0,0101	0,0091	0,0082	0,0074	0,0067
1	0,0679	0,0630	0,0583	0,0540	0,0500	0,0462	0,0427	0,0395	0,0365	0,0337
2	0,1393	0,1323	0,1254	0,1188	0,1125	0,1063	0,1005	0,0948	0,0894	0,0842
3	0,1904	0,1852	0,1798	0,1743	0,1687	0,1631	0,1574	0,1517	0,1460	0,1404
4	0,1951	0,1944	0,1933	0,1917	0,1898	0,1875	0,1849	0,1820	0,1789	0,1755
5	0,1600	0,1633	0,1662	0,1687	0,1708	0,1725	0,1738	0,1747	0,1753	0,1755
6	0,1093	0,1143	0,1191	0,1237	0,1281	0,1323	0,1362	0,1398	0,1432	0,1462
7	0,0640	0,0686	0,0732	0,0778	0,0824	0,0869	0,0914	0,0959	0,1022	0,1044
8	0,0328	0,0360	0,0393	0,0428	0,0463	0,0500	0,0537	0,0575	0,0614	0,0653
9	0,0150	0,0168	0,0188	0,0209	0,0232	0,0255	0,0280	0,0307	0,0334	0,0363
10	0,0061	0,0071	0,0081	0,0092	0,0104	0,0118	0,0132	0,0147	0,0164	0,0181
11	0,0023	0,0027	0,0032	0,0037	0,0043	0,0049	0,0056	0,0064	0,0073	0,0082
12	0,0008	0,0009	0,0011	0,0014	0,0016	0,0019	0,0022	0,0026	0,0030	0,0034
13	0,0002	0,0003	0,0004	0,0005	0,0006	0,0007	0,0008	0,0009	0,0011	0,0013
14	0,0001	0,0001	0,0001	0,0001	0,0002	0,0002	0,0003	0,0003	0,0004	0,0005
15	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	0,0001	0,0001	0,0001	0,0002
	5,1	**5,2**	**5,3**	**5,4**	**5,5**	**5,6**	**5,7**	**5,8**	**5,9**	**6,0**
0	0,0061	0,0055	0,0050	0,0045	0,0041	0,0037	0,0033	0,0030	0,0027	0,0025
1	0,0311	0,0287	0,0265	0,0244	0,0225	0,0207	0,0191	0,0176	0,0162	0,0149
2	0,0793	0,0746	0,0701	0,0659	0,0618	0,0580	0,0544	0,0509	0,0477	0,0446
3	0,1348	0,1293	0,1239	0,1185	0,1133	0,1082	0,1033	0,0985	0,0938	0,0892
4	0,1719	0,1681	0,1641	0,1600	0,1558	0,1515	0,1472	0,1428	0,1383	0,1339
5	0,1753	0,1748	0,1740	0,1728	0,1714	0,1697	0,1678	0,1656	0,1632	0,1606
6	0,1490	0,1515	0,1537	0,1555	0,1571	0,1584	0,1594	0,1601	0,1605	0,1606
7	0,1086	0,1125	0,1163	0,1200	0,1234	0,1267	0,1298	0,1326	0,1353	0,1377
8	0,0692	0,0731	0,0771	0,0810	0,0849	0,0887	0,0925	0,0962	0,0998	0,1033
9	0,0392	0,0423	0,0454	0,0486	0,0519	0,0552	0,0586	0,0620	0,0654	0,0688
10	0,0200	0,0220	0,0241	0,0262	0,0285	0,0309	0,0334	0,0359	0,0386	0,0413
11	0,0093	0,0104	0,0116	0,0129	0,0143	0,0157	0,0173	0,0190	0,0207	0,0225
12	0,0039	0,0045	0,0051	0,0058	0,0065	0,0073	0,0082	0,0092	0,0102	0,0113
13	0,0015	0,0018	0,0021	0,0024	0,0028	0,0032	0,0036	0,0041	0,0046	0,0052
14	0,0006	0,0007	0,0008	0,0009	0,0011	0,0013	0,0015	0,0017	0,0019	0,0022
15	0,0002	0,0002	0,0003	0,0003	0,0004	0,0005	0,0006	0,0007	0,0008	0,0009
16	0,0001	0,0001	0,0001	0,0001	0,0001	0,0002	0,0002	0,0002	0,0003	0,0003
17	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	0,0001	0,0001
	6,1	**6,2**	**6,3**	**6,4**	**6,5**	**6,6**	**6,7**	**6,8**	**6,9**	**7,0**
0	0,0022	0,0020	0,0018	0,0017	0,0015	0,0014	0,0012	0,0011	0,0010	0,0009
1	0,0137	0,0126	0,0116	0,0106	0,0098	0,0090	0,0082	0,0076	0,0070	0,0064
2	0,0417	0,0390	0,0364	0,0340	0,0318	0,0296	0,0276	0,0258	0,0240	0,0223
3	0,0848	0,0806	0,0765	0,0726	0,0688	0,0652	0,0617	0,0584	0,0552	0,0521
4	0,1294	0,1249	0,1205	0,1162	0,1118	0,1076	0,1034	0,0992	0,0952	0,0912
5	0,1579	0,1549	0,1519	0,1487	0,1454	0,1420	0,1385	0,1349	0,1314	0,1277
6	0,1605	0,1601	0,1595	0,1586	0,1575	0,1562	0,1546	0,1529	0,1511	0,1490
7	0,1399	0,1418	0,1435	0,1450	0,1462	0,1472	0,1480	0,1486	0,1489	0,1490
8	0,1066	0,1099	0,1130	0,1160	0,1188	0,1215	0,1240	0,1263	0,1284	0,1304
9	0,0723	0,0757	0,0791	0,0825	0,0858	0,0891	0,0923	0,0954	0,0985	0,1014
10	0,0441	0,0469	0,0498	0,0528	0,0558	0,0588	0,0618	0,0649	0,0679	0,0710
11	0,0245	0,0265	0,0285	0,0307	0,0330	0,0353	0,0377	0,0401	0,0426	0,0452
12	0,0124	0,0137	0,0150	0,0164	0,0179	0,0194	0,0210	0,0227	0,0245	0,0264
13	0,0058	0,0065	0,0073	0,0081	0,0089	0,0098	0,0108	0,0119	0,0130	0,0142
14	0,0025	0,0029	0,0033	0,0037	0,0041	0,0046	0,0052	0,0058	0,0064	0,0071
15	0,0010	0,0012	0,0014	0,0016	0,0018	0,0020	0,0023	0,0026	0,0029	0,0033
16	0,0004	0,0005	0,0005	0,0006	0,0007	0,0008	0,0010	0,0011	0,0013	0,0014
17	0,0001	0,0002	0,0002	0,0002	0,0003	0,0003	0,0004	0,0004	0,0005	0,0006
18	0,0000	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0002	0,0002	0,0002
19	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	0,0001

X	λ									
	7,1	**7,2**	**7,3**	**7,4**	**7,5**	**7,6**	**7,7**	**7,8**	**7,9**	**8,0**
0	0,0008	0,0007	0,0007	0,0006	0,0006	0,0005	0,0005	0,0004	0,0004	0,0003
1	0,0059	0,0054	0,0049	0,0045	0,0041	0,0038	0,0035	0,0032	0,0029	0,0027
2	0,0208	0,0194	0,0180	0,0167	0,0156	0,0145	0,0134	0,0125	0,0116	0,0107
3	0,0492	0,0464	0,0438	0,0413	0,0389	0,0366	0,0345	0,0324	0,0305	0,0286
4	0,0874	0,0836	0,0799	0,0764	0,0729	0,0696	0,0663	0,0632	0,0602	0,0573
5	0,1241	0,1204	0,1167	0,1130	0,1094	0,1057	0,1021	0,0986	0,0951	0,0916
6	0,1468	0,1445	0,1420	0,1394	0,1367	0,1339	0,1311	0,1282	0,1252	0,1221
7	0,1489	0,1486	0,1481	0,1474	0,1465	0,1454	0,1442	0,1428	0,1413	0,1396
8	0,1321	0,1337	0,1351	0,1363	0,1373	0,1382	0,1388	0,1392	0,1395	0,1396
9	0,1042	0,1070	0,1096	0,1121	0,1144	0,1167	0,1187	0,1207	0,1224	0,1241
10	0,0740	0,0770	0,0800	0,0829	0,0858	0,0887	0,0914	0,0941	0,0967	0,0993
11	0,0478	0,0504	0,0531	0,0558	0,0585	0,0613	0,0640	0,0667	0,0695	0,0722
12	0,0283	0,0303	0,0323	0,0344	0,0366	0,0388	0,0411	0,0434	0,0457	0,0481
13	0,0154	0,0168	0,0181	0,0196	0,0211	0,0227	0,0243	0,0260	0,0278	0,0296
14	0,0078	0,0086	0,0095	0,0104	0,0113	0,0123	0,0134	0,0145	0,0157	0,0169
15	0,0037	0,0041	0,0046	0,0051	0,0057	0,0062	0,0069	0,0075	0,0083	0,0090
16	0,0016	0,0019	0,0021	0,0024	0,0026	0,0030	0,0033	0,0037	0,0041	0,0045
17	0,0007	0,0008	0,0009	0,0010	0,0012	0,0013	0,0015	0,0017	0,0019	0,0021
18	0,0003	0,0003	0,0004	0,0004	0,0005	0,0006	0,0006	0,0007	0,0008	0,0009
19	0,0001	0,0001	0,0001	0,0002	0,0002	0,0002	0,0003	0,0003	0,0003	0,0004
20	0,0000	0,0000	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0002
21	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001
	8,1	**8,2**	**8,3**	**8,4**	**8,5**	**8,6**	**8,7**	**8,8**	**8,9**	**9,0**
0	0,0003	0,0003	0,0002	0,0002	0,0002	0,0002	0,0002	0,0002	0,0001	0,0001
1	0,0025	0,0023	0,0021	0,0019	0,0017	0,0016	0,0014	0,0013	0,0012	0,0011
2	0,0100	0,0092	0,0086	0,0079	0,0074	0,0068	0,0063	0,0058	0,0054	0,0050
3	0,0269	0,0252	0,0237	0,0222	0,0208	0,0195	0,0183	0,0171	0,0160	0,0150
4	0,0544	0,0517	0,0491	0,0466	0,0443	0,0420	0,0398	0,0377	0,0357	0,0337
5	0,0882	0,0849	0,0816	0,0784	0,0752	0,0722	0,0692	0,0663	0,0635	0,0607
6	0,1191	0,1160	0,1128	0,1097	0,1066	0,1034	0,1003	0,0972	0,0941	0,0911
7	0,1378	0,1358	0,1338	0,1317	0,1294	0,1271	0,1247	0,1222	0,1197	0,1171
8	0,1395	0,1392	0,1388	0,1382	0,1375	0,1366	0,1356	0,1344	0,1332	0,1318
9	0,1256	0,1269	0,1280	0,1290	0,1299	0,1306	0,1311	0,1315	0,1317	0,1318
10	0,1017	0,1040	0,1063	0,1084	0,1104	0,1123	0,1140	0,1157	0,1172	0,1186
11	0,0749	0,0776	0,0802	0,0828	0,0853	0,0878	0,0902	0,0925	0,0948	0,0970
12	0,0505	0,0530	0,0555	0,0579	0,0604	0,0629	0,0654	0,0679	0,0703	0,0728
13	0,0315	0,0334	0,0354	0,0374	0,0395	0,0416	0,0438	0,0459	0,0481	0,0504
14	0,0182	0,0196	0,0210	0,0225	0,0240	0,0256	0,0272	0,0289	0,0306	0,0324
15	0,0098	0,0107	0,0116	0,0126	0,0136	0,0147	0,0158	0,0169	0,0182	0,0194
16	0,0050	0,0055	0,0060	0,0066	0,0072	0,0079	0,0086	0,0093	0,0101	0,0109
17	0,0024	0,0026	0,0029	0,0033	0,0036	0,0040	0,0044	0,0048	0,0053	0,0058
18	0,0011	0,0012	0,0014	0,0015	0,0017	0,0019	0,0021	0,0024	0,0026	0,0029
19	0,0005	0,0005	0,0006	0,0007	0,0008	0,0009	0,0010	0,0011	0,0012	0,0014
20	0,0002	0,0002	0,0002	0,0003	0,0003	0,0004	0,0004	0,0005	0,0005	0,0006
21	0,0001	0,0001	0,0001	0,0001	0,0001	0,0002	0,0002	0,0002	0,0002	0,0003
22	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	1,0000	0,0001	0,0001	0,0001

X	λ									
	9,1	**9,2**	**9,3**	**9,4**	**9,5**	**9,6**	**9,7**	**9,8**	**9,9**	**10,0**
0	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0000
1	0,0010	0,0009	0,0009	0,0008	0,0007	0,0007	0,0006	0,0005	0,0005	0,0005
2	0,0046	0,0043	0,0040	0,0037	0,0034	0,0031	0,0029	0,0027	0,0025	0,0023
3	0,0140	0,0131	0,0123	0,0115	0,0107	0,0100	0,0093	0,0087	0,0081	0,0076
4	0,0319	0,0302	0,0285	0,0269	0,0254	0,0240	0,0226	0,0213	0,0201	0,0189
5	0,0581	0,0555	0,0530	0,0506	0,0483	0,0460	0,0439	0,0418	0,0398	0,0378
6	0,0881	0,0851	0,0822	0,0793	0,0764	0,0736	0,0709	0,0682	0,0656	0,0631
7	0,1145	0,1118	0,1091	0,1064	0,1037	0,1010	0,0982	0,0955	0,0928	0,0901
8	0,1302	0,1286	0,1269	0,1251	0,1232	0,1212	0,1191	0,1170	0,1148	0,1126
9	0,1317	0,1315	0,1311	0,1306	0,1300	0,1293	0,1284	0,1274	0,1263	0,1251
10	0,1198	0,1210	0,1219	0,1228	0,1235	0,1241	0,1245	0,1249	0,1250	0,1251
11	0,0991	0,1012	0,1031	0,1049	0,1067	0,1083	0,1098	0,1112	0,1125	0,1137
12	0,0752	0,0776	0,0799	0,0822	0,0844	0,0866	0,0888	0,0908	0,0928	0,0948
13	0,0526	0,0549	0,0572	0,0594	0,0617	0,0640	0,0662	0,0685	0,0707	0,0729
14	0,0342	0,0361	0,0380	0,0399	0,0419	0,0439	0,0459	0,0479	0,0500	0,0521
15	0,0208	0,0221	0,0235	0,0250	0,0265	0,0281	0,0297	0,0313	0,0330	0,0347
16	0,0118	0,0127	0,0137	0,0147	0,0157	0,0168	0,0180	0,0192	0,0204	0,0217
17	0,0063	0,0069	0,0075	0,0081	0,0088	0,0095	0,0103	0,0111	0,0119	0,0128
18	0,0032	0,0035	0,0039	0,0042	0,0046	0,0051	0,0055	0,0060	0,0065	0,0071
19	0,0015	0,0017	0,0019	0,0021	0,0023	0,0026	0,0028	0,0031	0,0034	0,0037
20	0,0007	0,0008	0,0009	0,0010	0,0011	0,0012	0,0014	0,0015	0,0017	0,0019
21	0,0003	0,0003	0,0004	0,0004	0,0005	0,0006	0,0006	0,0007	0,0008	0,0009
22	0,0001	0,0001	0,0002	0,0002	0,0002	0,0002	0,0003	0,0003	0,0004	0,0004
23	0,0000	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0002	0,0002
24	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	0,0001

X	λ				
	11	**12**	**13**	**14**	**15**
0	0,0000	0,0000	0,0000	0,0000	0,0000
1	0,0002	0,0001	0,0000	0,0000	0,0000
2	0,0010	0,0004	0,0002	0,0001	0,0000
3	0,0037	0,0018	0,0008	0,0004	0,0002
4	0,0102	0,0053	0,0027	0,0013	0,0006
5	0,0224	0,0127	0,0070	0,0037	0,0019
6	0,0411	0,0255	0,0152	0,0087	0,0048
7	0,0646	0,0437	0,0281	0,0174	0,0104
8	0,0888	0,0655	0,0457	0,0304	0,0194
9	0,1085	0,0874	0,0661	0,0473	0,0324
10	0,1194	0,1048	0,0859	0,0663	0,0486
11	0,1194	0,1144	0,1015	0,0844	0,0663
12	0,1094	0,1144	0,1099	0,0984	0,0829
13	0,0926	0,1056	0,1099	0,1060	0,0956
14	0,0728	0,0905	0,1021	0,1060	0,1024
15	0,0534	0,0724	0,0885	0,0989	0,1024
16	0,0367	0,0543	0,0719	0,0866	0,0960
17	0,0237	0,0383	0,0550	0,0713	0,0847
18	0,0145	0,0256	0,0397	0,0554	0,0706
19	0,0084	0,0161	0,0272	0,0409	0,0557
20	0,0046	0,0097	0,0177	0,0286	0,0418
21	0,0024	0,0055	0,0109	0,0191	0,0299
22	0,0012	0,0030	0,0065	0,0121	0,0204
23	0,0006	0,0016	0,0037	0,0074	0,0133
24	0,0003	0,0008	0,0020	0,0043	0,0083

25	0,0001	0,0004	0,0010	0,0024	0,0050
26	0,0000	0,0002	0,0005	0,0013	0,0029
27	0,0000	0,0001	0,0002	0,0007	0,0016
28	0,0000	0,0000	0,0001	0,0003	0,0009
29	0,0000	0,0000	0,0001	0,0002	0,0004
30	0,0000	0,0000	0,0000	0,0001	0,0002
31	0,0000	0,0000	0,0000	0,0000	0,0001
32	0,0000	0,0000	0,0000	0,0000	0,0001
33	0,0000	0,0000	0,0000	0,0000	0,0000
34	0,0000	0,0000	0,0000	0,0000	0,0000
35	0,0000	0,0000	0,0000	0,0000	0,0000
36	0,0000	0,0000	0,0000	0,0000	0,0000
37	0,0000	0,0000	0,0000	0,0000	0,0000
38	0,0000	0,0000	0,0000	0,0000	0,0000
39	0,0000	0,0000	0,0000	0,0000	0,0000

X			λ		
	16	**17**	**18**	**19**	**20**
0	0,0000	0,0000	0,0000	0,0000	0,0000
1	0,0000	0,0000	0,0000	0,0000	0,0000
2	0,0000	0,0000	0,0000	0,0000	0,0000
3	0,0001	0,0000	0,0000	0,0000	0,0000
4	0,0003	0,0001	0,0001	0,0000	0,0000
5	0,0010	0,0005	0,0002	0,0001	0,0001
6	0,0026	0,0014	0,0007	0,0004	0,0002
7	0,0060	0,0034	0,0018	0,0010	0,0005
8	0,0120	0,0072	0,0042	0,0024	0,0013
9	0,0213	0,0135	0,0083	0,0050	0,0029
10	0,0341	0,0230	0,0150	0,0095	0,0058
11	0,0496	0,0355	0,0245	0,0164	0,0106
12	0,0661	0,0504	0,0368	0,0259	0,0176
13	0,0814	0,0658	0,0509	0,0378	0,0271
14	0,0930	0,0800	0,0655	0,0514	0,0387
15	0,0992	0,0906	0,0786	0,0650	0,0516
16	0,0992	0,0963	0,0884	0,0772	0,0646
17	0,0934	0,0963	0,0936	0,0863	0,0760
18	0,0830	0,0909	0,0936	0,0911	0,0844
19	0,0699	0,0814	0,0887	0,0911	0,0888
20	0,0559	0,0692	0,0798	0,0866	0,0888
21	0,0426	0,0560	0,0684	0,0783	0,0846
22	0,0310	0,0433	0,0560	0,0676	0,0769
23	0,0216	0,0320	0,0438	0,0559	0,0669
24	0,0144	0,0226	0,0328	0,0442	0,0557
25	0,0092	0,0154	0,0237	0,0336	0,0446
26	0,0057	0,0101	0,0164	0,0246	0,0343
27	0,0034	0,0063	0,0109	0,0173	0,0254
28	0,0019	0,0038	0,0070	0,0117	0,0181
29	0,0011	0,0023	0,0044	0,0077	0,0125
30	0,0006	0,0013	0,0026	0,0049	0,0083
31	0,0003	0,0007	0,0015	0,0030	0,0054
32	0,0001	0,0004	0,0009	0,0018	0,0034
33	0,0001	0,0002	0,0005	0,0010	0,0020
34	0,0000	0,0001	0,0002	0,0006	0,0012
35	0,0000	0,0000	0,0001	0,0003	0,0007
36	0,0000	0,0000	0,0001	0,0002	0,0004
37	0,0000	0,0000	0,0000	0,0001	0,0002
38	0,0000	0,0000	0,0000	0,0000	0,0001
39	0,0000	0,0000	0,0000	0,0000	0,0001

Tabel 4 Standaardnormale kansen

Het tabelelement bij Z is de kans P(Z ≥ z)

z	0	1	2	3	4	5	6	7	8	9
0,0	0,500	0,496	0,492	0,488	0,484	0,480	0,476	0,472	0,468	0,464
0,1	0,460	0,456	0,452	0,448	0,444	0,440	0,436	0,433	0,429	0,425
0,2	0,421	0,417	0,413	0,409	0,405	0,401	0,397	0,394	0,390	0,386
0,3	0,382	0,378	0,374	0,371	0,367	0,363	0,359	0,356	0,352	0,348
0,4	0,345	0,341	0,337	0,334	0,330	0,326	0,323	0,319	0,316	0,312
0,5	0,309	0,305	0,302	0,298	0,295	0,291	0,288	0,284	0,281	0,278
0,6	0,274	0,271	0,268	0,264	0,261	0,258	0,255	0,251	0,248	0,245
0,7	0,242	0,239	0,236	0,233	0,230	0,227	0,224	0,221	0,218	0,215
0,8	0,212	0,209	0,206	0,203	0,200	0,198	0,195	0,192	0,189	0,187
0,9	0,184	0,181	0,179	0,176	0,174	0,171	0,169	0,166	0,164	0,161
1,0	0,159	0,156	0,154	0,152	0,149	0,147	0,145	0,142	0,140	0,138
1,1	0,136	0,133	0,131	0,129	0,127	0,125	0,123	0,121	0,119	0,117
1,2	0,115	0,113	0,111	0,109	0,107	0,106	0,104	0,102	0,100	0,099
1,3	0,097	0,095	0,093	0,092	0,090	0,089	0,087	0,085	0,084	0,082
1,4	0,081	0,079	0,078	0,076	0,075	0,074	0,072	0,071	0,069	0,068
1,5	0,067	0,066	0,064	0,063	0,062	0,061	0,059	0,058	0,057	0,056
1,6	0,055	0,054	0,053	0,052	0,051	0,049	0,048	0,047	0,046	0,046
1,7	0,045	0,044	0,043	0,042	0,041	0,040	0,039	0,038	0,038	0,037
1,8	0,036	0,035	0,034	0,034	0,033	0,032	0,031	0,031	0,030	0,029
1,9	0,029	0,028	0,027	0,027	0,026	0,026	0,025	0,024	0,024	0,023
2,0	0,023	0,022	0,022	0,021	0,021	0,020	0,020	0,019	0,019	0,018
2,1	0,018	0,017	0,017	0,017	0,016	0,016	0,015	0,015	0,015	0,014
2,2	0,014	0,014	0,013	0,013	0,013	0,012	0,012	0,012	0,011	0,011
2,3	0,011	0,010	0,010	0,010	0,010	0,009	0,009	0,009	0,009	0,008
2,4	0,008	0,008	0,008	0,008	0,007	0,007	0,007	0,007	0,007	0,006
2,5	0,006	0,006	0,006	0,006	0,006	0,005	0,005	0,005	0,005	0,005
2,6	0,005	0,005	0,004	0,004	0,004	0,004	0,004	0,004	0,004	0,004
2,7	0,003	0,003	0,003	0,003	0,003	0,003	0,003	0,003	0,003	0,003
2,8	0,003	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,002
2,9	0,002	0,002	0,002	0,002	0,002	0,002	0,002	0,001	0,001	0,001

Tabel 5 t-kritieke waarden

met overschrijdingskans rechts van t

d.f	$t_{0,25}$	$t_{0,10}$	$t_{0,05}$	$t_{0,025}$	$t_{0,010}$	$t_{0,005}$	$t_{0,0025}$	$t_{0,0010}$	$t_{0,0005}$
1	1,00	3,08	6,31	12,70	31,80	63,70	127	318	637
2	0,82	1,89	2,92	4,30	6,96	9,92	14,10	22,30	31,60
3	0,76	1,64	2,35	3,18	4,54	5,84	7,45	10,20	12,90
4	0,74	1,53	2,13	2,78	3,75	4,60	5,60	7,17	8,61
5	0,73	1,48	2,02	2,57	3,36	4,03	4,77	5,89	6,87
6	0,72	1,44	1,94	2,45	3,14	3,71	4,32	5,21	5,96
7	0,71	1,41	1,89	2,36	3,00	3,50	4,03	4,79	5,41
8	0,71	1,40	1,86	2,31	2,90	3,36	3,83	4,50	5,04
9	0,70	1,38	1,83	2,26	2,82	3,25	3,69	4,30	4,78
10	0,70	1,37	1,81	2,23	2,76	3,17	3,58	4,14	4,59
11	0,70	1,36	1,80	2,20	2,72	3,11	3,50	4,02	4,44
12	0,70	1,36	1,78	2,18	2,68	3,05	3,43	3,93	4,32
13	0,69	1,35	1,77	2,16	2,65	3,01	3,37	3,85	4,22
14	0,69	1,35	1,76	2,14	2,62	2,98	3,33	3,79	4,14
15	0,69	1,34	1,75	2,13	2,60	2,95	3,29	3,73	4,07
16	0,69	1,34	1,75	2,12	2,58	2,92	3,25	3,69	4,01
17	0,69	1,33	1,74	2,11	2,57	2,90	3,22	3,65	3,97
18	0,69	1,33	1,73	2,10	2,55	2,88	3,20	3,61	3,92
19	0,69	1,33	1,73	2,09	2,54	2,86	3,17	3,58	3,88
20	0,69	1,33	1,72	2,09	2,53	2,85	3,15	3,55	3,85
21	0,69	1,32	1,72	2,08	2,52	2,83	3,14	3,53	3,82
22	0,69	1,32	1,72	2,07	2,51	2,82	3,12	3,50	3,79
23	0,69	1,32	1,71	2,07	2,50	2,81	3,10	3,48	3,77
24	0,68	1,32	1,71	2,06	2,49	2,80	3,09	3,47	3,75
25	0,88	1,32	1,71	2,06	2,49	2,79	3,08	3,45	3,73
26	0,68	1,31	1,71	2,06	2,48	2,78	3,07	3,43	3,71
27	0,68	1,31	1,70	2,05	2,47	2,77	3,06	3,42	3,69
28	0,68	1,31	1,70	2,05	2,47	2,76	3,05	3,41	3,67
29	0,68	1,31	1,70	2,05	2,46	2,76	3,04	3,40	3,66
30	0,68	1,31	1,70	2,04	2,46	2,75	3,03	3,39	3,65
40	0,68	1,30	1,68	2,02	2,42	2,70	2,97	3,31	3,55
60	0,68	1,30	1,67	2,00	2,39	2,66	2,92	3,23	3,46
120	0,68	1,29	1,66	1,98	2,36	2,62	2,86	3,16	3,37

Tabel 6 χ^2- kritieke waarden

met overschrijdingskans rechts van χ^2

d.f.	$\chi^2_{.25}$	$\chi^2_{.10}$	$\chi^2_{.05}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$	$\chi^2_{.001}$
1	1,32	2,71	3,84	5,02	6,63	7,88	10,8
2	2,77	4,61	5,99	7,38	9,21	10,6	13,8
3	4,11	6,25	7,81	9,35	11,3	12,8	16,3
4	5,39	7,78	9,49	11,3	13,3	14,9	18,5
5	6,63	9,24	11,1	12,8	15,1	16,7	20,5
6	7,84	10,6	12,6	14,4	16,8	18,5	22,5
7	9,04	12,0	14,1	16,0	18,5	20,3	24,3
8	10,2	13,4	15,5	17,5	20,1	22,0	26,1
9	11,4	14,7	16,9	19,0	21,7	23,6	27,9
10	12,5	16,0	18,3	20,5	23,2	25,2	29,6
11	13,7	17,3	19,7	21,9	24,7	26,8	31,3
12	14,8	18,5	21,0	23,3	26,2	28,3	32,9
13	16,0	19,8	22,4	24,7	27,7	29,8	34,5
14	17,1	21,1	23,7	26,1	29,1	31,3	36,1
15	18,2	22,3	25,0	27,5	30,6	32,8	37,7
16	19,4	23,5	26,3	28,8	32,0	34,3	39,3
17	20,5	24,8	27,6	30,2	33,4	35,7	40,8
18	21,6	26,0	28,9	31,5	34,8	37,2	42,3
19	22,7	27,2	30,1	32,9	36,2	38,6	32,8
20	23,8	28,4	31,4	34,2	37,6	40,0	45,3
21	24,9	29,6	32,7	35,5	38,9	41,4	46,8
22	26,0	30,8	33,9	36,8	40,3	42,8	48,3
23	27,1	32,0	35,2	38,1	41,6	44,2	49,7
24	28,2	33,2	36,4	39,4	32,0	45,6	51,2
25	29,3	34,4	37,7	40,6	44,3	46,9	52,6
26	30,4	35,6	38,9	41,9	45,6	48,3	54,1
27	31,5	36,7	40,1	43,2	47,0	49,6	55,5
28	32,6	37,9	41,3	44,5	48,3	51,0	56,9
29	33,7	39,1	42,6	45,7	49,6	52,3	58,3
30	34,8	40,3	43,8	47,0	50,9	53,7	59,7
40	45,6	51,8	55,8	59,3	63,7	66,8	73,4
50	56,3	63,2	67,5	71,4	76,2	79,5	86,7
60	67,0	74,4	79,1	83,3	88,4	92,0	99,6
70	77,6	85,5	90,5	95	100	104	112
80	88,1	96,6	102	107	112	116	125
90	98,6	108	113	118	124	128	137
100	109	118	124	130	136	140	149

Tabel 7 F-kritieke waarden

met overschrijdingskans rechts van F

Vrijheidsgraden in de noemer	Vrijheidsgraden in de teller											
		1	2	3	4	5	6	8	10	20	40	∞
1	F0,25	5,83	7,50	8,20	8,58	8,82	8,98	9,19	9,32	9,58	9,71	9,85
	F0,1	39,90	49,50	53,60	55,80	57,20	58,20	59,40	60,20	61,70	62,50	63,30
	F0,05	161	200	216	225	230	224	239	242	248	251	254
2	F0,25	2,57	3,00	3,15	3,23	3,28	3,31	3,35	3,38	3,43	3,45	3,48
	F0,1	8,53	9,00	9,16	9,24	9,29	9,33	9,37	9,39	9,44	9,47	9,49
	F0,05	18,50	19,00	19,20	19,20	19,30	19,30	19,40	19,40	19,40	19,50	19,50
	F0,01	98,50	99,00	99,20	99,20	99,30	99,30	99,40	99,40	99,40	99,50	99,50
	F0,001	998	999	999	999	999	999	999	999	999	999	999
3	F0,25	2,02	2,28	2,36	2,39	2,41	2,42	2,44	2,44	2,46	2,47	2,47
	F0,1	5,54	5,46	5,39	5,34	5,31	5,28	5,25	5,23	5,18	5,16	5,13
	F0,05	10,10	9,55	9,28	9,12	9,10	8,94	8,85	8,79	8,66	8,59	8,53
	F0,01	34,10	30,80	29,50	28,70	28,20	27,90	27,50	27,20	26,70	26,40	26,10
	F0,001	167	149	141	137	135	133	131	129	126	125	124
4	F0,25	1,81	2,00	2,05	2,06	2,07	2,08	2,08	2,08	2,08	2,08	2,08
	F0,1	4,54	4,32	4,19	4,11	4,05	4,01	3,95	3,92	3,84	3,80	3,76
	F0,05	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,96	5,80	5,72	5,63
	F0,01	21,20	18,00	16,70	16,00	15,50	15,20	14,80	14,50	14,00	13,70	13,50
	F0,001	74,10	61,30	56,20	53,40	51,70	50,50	49,00	48,10	46,10	45,10	44,10
5	F0,25	1,69	1,85	1,88	1,89	1,89	1,89	1,89	1,89	1,88	1,88	1,87
	F0,1	4,06	3,78	3,62	3,52	3,45	3,40	3,34	3,30	3,21	3,16	3,10
	F0,05	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,74	4,56	4,46	4,36
	F0,01	16,30	13,30	12,10	11,40	11,00	10,70	10,30	10,10	9,55	9,29	9,02
	F0,001	47,20	37,10	33,20	31,10	29,80	28,80	27,60	26,90	25,40	24,60	23,80
6	F0,25	1,62	1,76	1,78	1,79	1,79	1,78	1,77	1,77	1,76	1,75	1,74
	F0,1	3,78	3,46	3,29	3,18	3,11	3,05	2,98	2,94	2,84	2,78	2,72
	F0,05	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,06	3,87	3,77	3,67
	F0,01	13,70	10,90	9,78	9,15	8,75	8,47	8,10	7,87	7,40	7,14	6,88
	F0,001	35,50	27,00	23,70	21,90	20,80	20,00	19,00	18,40	17,10	16,40	15,80
7	F0,25	1,57	1,70	1,72	1,72	1,71	1,71	1,70	1,69	1,67	1,66	1,65
	F0,1	3,59	3,26	3,07	2,96	2,88	2,83	2,75	2,70	2,59	2,54	2,47
	F0,05	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,64	3,44	3,34	3,23
	F0,01	12,20	9,55	8,45	7,85	7,46	7,19	6,84	6,62	6,16	5,91	5,65
	F0,001	29,30	21,70	18,80	17,20	16,20	15,50	14,60	14,10	12,90	12,30	11,70
8	F0,25	1,54	1,66	1,67	1,66	1,66	1,65	1,64	1,63	1,61	1,59	1,58
	F0,1	3,46	3,11	2,92	2,81	2,73	2,67	2,59	2,54	2,42	2,36	2,29
	F0,05	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,35	3,15	3,04	2,93
	F0,01	11,30	8,65	7,59	7,01	6,63	6,37	6,03	5,81	5,36	5,12	4,86
	F0,001	25,40	18,50	15,80	14,40	13,50	12,90	12,00	11,50	10,50	9,92	9,33
9	F0,25	1,51	1,62	1,63	1,63	1,62	1,61	1,60	1,59	1,56	1,55	1,53
	F0,1	3,36	3,01	2,81	2,69	2,61	2,55	2,47	2,42	2,30	2,23	2,16
	F0,05	5,12	4,26	3,86	3,63	3,48	3,37	3,23	3,14	2,94	2,83	2,71
	F0,01	10,60	8,02	6,99	6,42	6,06	5,80	5,47	5,26	4,81	4,57	4,31
	F0,001	22,90	16,40	13,90	12,60	11,70	11,10	10,40	9,89	8,90	8,37	7,81
10	F0,25	1,49	1,60	1,60	1,59	1,59	1,58	1,56	1,55	1,52	1,51	1,48
	F0,1	3,28	2,92	2,73	2,61	2,52	2,46	2,38	2,32	2,20	2,13	2,06
	F0,05	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,98	2,77	2,66	2,54
	F0,01	10,00	7,56	6,55	5,99	5,64	3,39	5,06	4,85	4,41	4,17	3,91
	F0,001	21,00	14,90	12,60	11,30	10,50	9,92	9,20	8,75	7,80	7,30	6,76

Vrijheidsgraden in de noemer		Vrijheidsgraden in de teller										
		1	2	3	4	5	6	8	10	20	40	∞
12	F0,25	1,56	1,56	1,56	1,55	1,54	1,53	1,51	1,50	1,47	1,45	1,42
	F0,1	3,18	2,81	2,61	2,48	2,39	2,33	2,24	2,19	2,06	1,99	1,90
	F0,05	4,75	3,89	3,49	3,26	3,11	3,00	2,85	2,75	2,54	2,43	2,30
	F0,01	9,33	6,93	5,95	5,41	5,06	4,82	4,50	4,30	3,86	3,62	3,36
	F0,001	18,60	13,00	10,80	9,63	8,89	8,38	7,71	7,29	6,40	5,93	5,42
14	F0,25	1,44	1,53	1,53	1,52	1,51	1,50	1,48	1,46	1,43	1,41	1,38
	F0,1	3,10	2,73	2,52	2,39	2,31	2,24	2,15	2,10	1,96	1,89	1,80
	F0,05	4,60	3,74	3,34	3,11	2,96	2,85	2,70	2,60	2,39	2,27	2,13
	F0,01	8,86	5,51	5,56	5,04	4,69	4,46	4,14	3,94	3,51	3,27	3,00
	F0,001	17,10	11,80	9,73	8,62	7,92	7,43	6,80	6,40	5,56	5,10	4,60
16	F0,25	1,42	1,51	1,51	1,50	1,48	1,48	1,46	1,45	1,40	1,37	1,34
	F0,1	3,05	2,67	2,46	2,33	2,24	2,18	2,09	2,03	1,89	1,81	1,72
	F0,05	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,49	2,28	2,15	2,01
	F0,01	8,53	6,23	5,29	4,77	4,44	4,20	3,89	3,69	3,26	3,02	2,75
	F0,001	16,10	11,00	9,00	7,94	7,27	6,81	6,19	5,81	4,99	4,54	4,06
20	F0,25	1,40	1,49	1,48	1,46	1,45	1,44	1,42	1,40	1,36	1,33	1,29
	F0,1	2,97	2,59	2,38	2,25	2,16	2,09	2,00	1,94	1,79	1,71	1,61
	F0,05	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,35	2,12	1,99	1,84
	F0,01	8,10	5,85	4,94	4,43	4,10	3,87	3,56	3,37	2,94	2,69	2,42
	F0,001	14,80	9,95	8,10	7,10	6,46	6,02	5,44	5,08	4,29	3,86	3,38
30	F0,25	1,38	1,45	1,44	1,42	1,41	1,39	1,37	1,35	1,30	1,27	1,23
	F0,1	2,88	2,49	2,28	2,14	2,05	1,98	1,88	1,82	1,67	1,57	1,46
	F0,05	4,17	3,32	2,92	2,69	2,53	2,42	2,27	2,16	1,93	1,79	1,62
	F0,01	7,56	5,39	4,51	4,02	3,70	3,47	3,17	2,98	2,55	2,30	2,01
	F0,001	13,30	8,77	7,05	6,12	5,53	3,12	4,58	4,24	3,49	3,07	2,59
40	F0,25	1,36	1,44	1,42	1,40	1,39	1,37	1,35	1,33	1,28	1,24	1,19
	F0,1	2,84	2,44	2,23	2,09	2,00	1,93	1,83	1,76	1,61	1,51	1,38
	F0,05	4,08	3,23	2,84	2,61	2,45	2,34	2,18	2,08	1,84	1,69	1,51
	F0,01	7,31	5,18	4,31	3,83	3,51	3,29	2,99	2,80	2,37	2,11	1,80
	F0,001	12,60	8,25	6,60	5,70	5,13	4,73	4,21	3,87	3,15	2,73	2,23
60	F0,25	1,35	1,42	1,41	1,38	1,37	1,35	1,32	1,30	1,25	1,21	1,15
	F0,1	2,79	2,39	2,18	2,04	1,95	1,87	1,77	1,71	1,54	1,44	1,29
	F0,05	4,00	3,15	2,76	2,53	2,37	2,25	2,10	1,99	1,75	1,59	1,39
	F0,01	7,08	4,98	4,13	3,65	3,34	3,12	2,82	2,63	2,20	1,94	1,50
	F0,001	12,00	7,76	6,17	5,31	4,76	4,37	3,87	3,54	2,83	2,41	1,89
120	F0,25	1,34	1,40	1,39	1,37	1,35	1,33	1,30	1,28	1,22	1,18	1,10
	F0,1	2,75	2,35	2,13	1,99	1,90	1,82	1,72	1,65	1,48	1,37	1,19
	F0,05	3,92	3,07	2,68	2,45	2,29	2,17	2,02	1,91	1,66	1,50	1,25
	F0,01	6,85	4,79	3,95	3,48	3,17	2,96	2,66	2,47	2,03	1,76	1,38
	F0,001	11,40	7,32	5,79	4,95	4,42	4,04	3,55	3,24	2,53	2,11	1,54
∞	F0,25	1,32	1,39	1,37	1,35	1,33	1,31	1,28	1,25	1,19	1,14	1,00
	F0,1	2,71	2,30	2,08	1,94	1,85	1,77	1,67	1,60	1,42	1,30	1,00
	F0,05	3,84	3,00	2,60	2,37	2,21	2,10	1,94	1,83	1,57	1,39	1,00
	F0,01	6,63	4,61	3,78	3,32	3,02	2,80	2,51	2,32	1,88	1,59	1,00
	F0,001	10,80	6,91	5,42	4,62	4,10	3,74	3,27	2,96	2,27	1,84	1,00

Bijlage 5 Vakliteratuur

Berenson, M. L. & Levine D. M. (1996). *Basic Business Statistics: Concepts and Applications*. Englewood Cliffs, NJ, Prentice Hall.

Buglear, J. (2005). *Quantitative methods for Business*. Elsevier Butterworth Heinemann.

Buhrman, J. M. (1996). *Basisboek statistiek*. Wolters-Noordhoff.

Buys, A. (1998). *Statistiek om mee te werken*. EPN.

Carette, I. & Feryn-Patfoort, C. (1990). *Statistiek in modules*. Wolters Leuven.

Hogg, R. V. & Tanis, E. A. (1997). *Probability and Statistical Inference*. Englewood Cliffs, NJ, Prentice Hall.

Levin, R. I. & Rubin, D. S. (1998). *Statistics for Management*. Prentice Hall International.

Maryns, J. & Paemeleire, R. (1978). *Statistiek, bedrijfseconomisch benaderd*. MIM.

McClave, Benson en Sincich (2007). *Statistiek: een inleiding voor het hoger onderwijs*. Pearson.

Moore, D. S. & McCabe, G. P. (1994). *Statistiek in de praktijk*. Academic Service.

Piessens, B. (1975). *Waarschijnlijkheidsrekening en Statistiek*. Wetenschappelijke uitgeverij E. Story-Scientia.

Sharpe, N., De Veaux, R. & Velleman, P. (2015). *Business Statistics*. Pearson.

Spiegel, M. R., Schiller J. J. & Srinivasan Alu R. (2000). *Probability and Statistics*. Mc Graw-Hill Book Company.

Spiegel, M.R. (1961). *Statistics*. McGraw-Hill Book Company.

Touw, P. & Hoogduin, L. (2002). *Statistiek voor accountancy*. Academic Service.

Van den Broeck, E. & Cuypers, E. (2013). *MS Excel 2013*. Uitgeverij De Boeck.

Van den Broeck, E. & Cuypers, E. (2016). *MS Excel 2016*. Uitgeverij De Boeck.

Van Der Elst, J. (2005). *Statistiek in de Bedrijfskunde*. Uitgeverij Elbis.

Wonnacott, T. H. & Wonnacott, R. J. (1984). *Introductory Statistics for Business and Economics*. John Wiley & Sons.